GENETIC MECHANISMS CONTROLLING HUMAN HAIR GROWTH

Gina M. DeStefano

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Genetic Mechanisms Controlling Human Hair Growth

Gina M. DeStefano

The genetic underpinnings of human hair growth are complex, relying upon several mechanisms to regulate gene expression. As such, inherited skin and hair disorders can arise from a variety of mutational events in the genome, from changes in single nucleotides to structural rearrangements of chromosomes. Importantly, inherited conditions affecting hair growth can be used as models to interrogate the molecular basis of the disease, and obtain novel insight into mechanisms and pathways required for normal hair growth. The approaches used to identify pathogenic mutations in a given disorder will depend on the inheritance pattern and disease prevalence in the population. In rare, Mendelian disorders of the skin and hair, the genetic architecture is often composed of rare variants with large effects, detected by linkage or whole-exome/genome sequencing. In contrast, polygenic disorders are composed of common and rare variants that contribute small to moderate effects and can be detected using Genome-Wide Association Studies (GWAS).

The primary goal of my thesis research is to identify and characterize genetic mechanisms controlling human hair growth. To accomplish this, I have studied three inherited conditions affecting human hair growth as genetic models in which I performed detailed functional and molecular analyses of the causal genetic lesions and their downstream effects on gene expression in the hair follicle. My thesis work revolved around the use of three major approaches to identify and characterize genetic

mechanisms underlying human hair growth: 1. Identifying genomic effects on human hair growth in a rare, sporadic Mendelian disorder (Chapter II), 2. Characterizing single-gene effects on human hair growth in a rare, familial condition (Chapter III), and 3. Functional analysis of rare, non-coding variants in a complex polygenic autoimmune disease, alopecia areata (Chapter IV).

In the first part of my thesis work, I investigated the genetic mechanism associated with X-linked hypertrichosis (XLH), a very rare condition of excessive hair overgrowth. We identified a large interchromosomal insertion at chrXq27.1 that completely cosegregated with the phenotype, and was consistent with findings of interchromosomal insertions in two previously reported XLH families. Remarkably, the insertions in all three families occur at the exact same palindromic sequence, and because the sequences contained within each insertion are distinct, we hypothesized that the presence of the insertion (rather than its content) may be responsible for the excessive hair overgrowth phenotype. I then tested the impact of the insertion on the expression of the surrounding genes and found that *FGF13* levels were selectively and dramatically reduced in patient hair follicles, suggesting a position effect as a result of the interchromosomal insertion. We postulate that the presence of this insertion disrupts key inter- and intrachromosomal interactions required for normal hair growth.

In the second part of my thesis, I identified single-gene mutations that affect human hair growth by investigating the genetic basis of autosomal recessive congenital generalized hypertrichosis terminalis (CGHT) in a consanguineous family. We performed whole-exome sequencing and identified a novel, rare splice variant in *ABCA5* that cosegregates with the phenotype in a homozygous recessive manner. I found that

ABCA5 is highly expressed in human skin and hair follicles, and its expression pattern is conserved in mouse tissues as well. The *ABCA5* mutation in CGHT leads to a complete loss-of-function in patient hair follicles, as well as reduced lysosome function and cholesterol transport, a finding consistent with defects in *Abca5*$^{-/-}$ mice. Moreover, we identified a deletion spanning *ABCA5* in an unrelated sporadic CGHT case and found that *ABCA5* levels were dramatically reduced in patient hair follicles. Collectively, our findings point to a novel role for ABCA5 in regulating hair growth.

In the third part of my thesis, I characterized the genetic architecture of a complex, polygenic disease affecting hair growth by studying rare, non-coding variants in alopecia areata. The Christiano lab previously performed the first GWAS, identifying the *ULBP3/6* locus on chr.6q25.1 encoding NKG2D T cell receptor ligands as the most significant association outside the *HLA* region. A strong upregulation of these ligands was observed on both human and mouse hair follicles, and we recently showed that T cells bearing the NKG2D receptor are both necessary and sufficient to induce disease in the mouse model. To identify the susceptibility variants at *ULBP3/6* in human AA, we performed targeted deep resequencing and functional genomics studies and identified three rare, novel variants that reside within *ULBP6* regulatory elements and CTCF binding sites. I found that these variants disrupt CTCF binding and regulatory activity *in vitro,* and CTCF binding is enriched at *ULBP6 in vivo.* Future studies examining long-range reporter activity and CTCF-mediated interactions will define the role of CTCF in the repression of the *ULBP3/6* genes in the human hair follicle.

Collectively, I used a variety of genetic approaches to identify novel genes and mechanisms controlling human hair growth.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

Chapter I

Chapter II

Chapter III

Chapter IV

LIST OF ABBREVIATIONS

Human genes and proteins are written in italicized and non-italicized capital letters, respectively. Mouse nomenclature is the same but only the first letter is capitalized.

3C: Chromosome Conformation Capture
AA: alopecia areata
ABC: ATP-Binding Cassette transporter
ABCA5: ATP-Binding Cassette, Sub-Family A (ABC1), Member 5
AD: autosomal dominant
ALDH2: aldehyde dehydrogenase family 2
APCDD1: adenomatosis polyposis coli down-regulated 1
AR: autosomal recessive
ARWH: autosomal recessive woolly hair
AS-C: achaete-scute complex (*Drosophila melanogaster*)
ATP: adenosine triphosphate
β2M: beta-2-microglobulin
BAC: bacterial artificial chromosome
BAF: bafilomycin A1
BIM: BCL2-Like 11, apoptosis facilitator (alias gene name: **BCL2L11**)
BMP: bone morphogenic protein
bp: base pair (s)
cDNA: complementary DNA
CeD: celiac disease
CGH: comparative genome hybridization
CGHT: Congenital generalized hypertrichosis terminalis
ChIA-PET: chromatin interactions analysis with paired-end tag sequencing
ChIP: chromatin immunoprecipitation
Chr: chromosome
CNV: copy number variation
CpG: cytosine-phosphate-guanine dinucleotide
CREB: cyclic AMP response element binding protein
CTCF: CCCTC-binding factor (zing finger protein)
Cxorf66: chromosome X open reading frame 66
DAAM: disheveled associated activator of morphogenesis
DAPI: 4',6-diamidino-2-phenylindole
dbSNP: single nucleotide polymorphism database
DGK: diacylglycerol (DAG) kinase
DKK: dickkopf homolog
DMEM: Dulbecco's modified eagle medium
DNA: deoxyribonucleic acid
DP: dermal papilla
DS: dermal sheath
DSG: desmoglein
DZ: dizygotic (twins)

E14.5: 14.5 embryonic days post coitus
EDC: epidermal differentiation complex
EDA: ectodysplasin-A
EDAR: ectodysplasin-A receptor
EGF: epidermal growth factor
EGR3: early growth response protein 3
ELK1: ETS domain containing protein 1
ENCODE: encyclopedia of DNA elements
ER: endoplasmic reticulum
eQTL: expression quantitative trait locus
FACS: fluorescence-activated cell sorting
FAIRE: formaldehyde assisted isolation of regulatory elements
FAM162A: family with sequence similarity 162, member A
FANTOM5: functional annotation of the mammalian genome 5
FB: fibroblast
FGF: fibroblast growth factor
FGF13: Fibroblast growth factor 13
FISH: fluorescence *in situ* hybridization
FOXN1: forkhead box N1
GAPDH: glyceraldehyde-3-phosphate dehydrogenase
GPC3: glypican-3
GWAS: genome-wide association study
H19: H19, imprinted maternally expressed transcript
H3K4me1: histone tail 3 lysine residue 4 monomethylation
H3K4me3: histone tail 3 lysine residue 4 trimethylation
H3K27Ac: histone tail 3, lysine residue 27, acetylated
HED: hypohidrotic ectodermal dysplasia
HEK: human embryonic kidney cells
HF: hair follicle
HHS: hereditary hypotrichosis simplex
Hi-C: chromatin conformation capture (3C) analysis with high-throughput sequencing
HLA: human leukocyte antigen
HOX: Homeobox
HS: hair shaft
hTERT: human telomerase reverse transcriptase, cell line
IBD: identical by descent
ICR: imprinting control region
IFNG: interferon gamma
IgG: immunoglobulin G
IGF: insulin-like growth factor
IL: interleukin
IRS: inner root sheath
ISH: *in situ* hybridization
IT: insertional translocation
JAK: Janus family protein tyrosine kinase
K: keratin protein

Kb: kilobase (s)
KC: keratinocyte
KDa: kilodalton
KIF6: kinesin family member 6
KRT: keratin gene
LAH: localized autosomal recessive hypotrichosis
LAMP: lysosome associated membrane protein
LC3: light chain 3 (alias gene name: **MAP1LC3:** microtubule-associated protein 1 light chain 3)
LD: linkage disequilibrium
LDLr: low density lipoprotein receptor
LIPH: lipase H
lincRNA: long, non-coding RNA
LOD: logarithm of the odds
LPA: lysophosphatidic acid
LPAR6: lysophosphatidic acid receptor 6 (alias gene name: P2RY5)
LTR: long terminal repeat
LRRC32: leucine rich repeat containing 32 (alias gene name: **GARP**: Glycoprotein A Repetitions Predominant)
MAPK: mitogen-activated protein kinase
Mb: megabase (s)
MCF7: Michigan Cancer Foundation-7, breast cancer cell line
MHC: major histocompatability complex
MICA/B: MHC class I chain-related gene A or B
miRNA: microRNA
mRNA: messenger RNA
Mx: matrix
MZ: monozygotic (twins)
NCBI: National Center for Biotechnology Information
NGS: next-generation sequencing
NK: natural killer cell
NKG2D: natural-killer group 2, member D (alias gene name: **KLRK1**: killer cell lectin-like receptor subfamily K, member 1)
OCT: optimal cutting temperature compound
OMIM: Online Mendelian Inheritance in Man
ORS: outer root sheath
P1: postnatal day 1
PBMC: peripheral blood mononuclear cell
PCR: polymerase chain reaction
PKC: protein kinase C
PMA: phorbol 12-myristate 13-acetate
Pol II: RNA polymerase II
PPAR: peroxisome proliferator-activated receptor
PXE: pseudoxanthoma elasticum
qRT-PCR: quantitative reverse-transcriptase PCR
RA: rheumatoid arthritis

RF: recombination fraction
RNA: ribonucleic acid
RNA-seq: RNA sequencing
RSPO: R-spondin
SCID: severe combined immunodeficiency
SD: standard deviation
SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE: super enhancer
SEM: standard error of the mean
SH2B3: SH2B adaptor protein 3 (gene alias name: **LNK:** Lymphocyte-Specific Adapter Protein)
SHH: sonic hedgehog
SNP: single nucleotide polymorphism
SNV: single nucleotide variant
SOMA: SNP oligonucleotide microarray analysis
SOX: SRY (sex determining region Y)-box
STAT: signal transducer and activator of transcription
STX17: syntaxin 17
SUZ12: suppressor of zeste 12 (*Drosophila melanogaster* homolog)
T1D: type I diabetes
TF: transcription factor
TFBS: transcription factor binding site
TGF: transforming growth factor
Th2: T helper cell type 2
TR: thyroid hormone receptor
TRPS1: trichorhinophalangeal syndrome I
UCSC: University of California, Santa Cruz
ULBP: cytomegalovirus UL16-binding protein
Var: variant
VEGF: vascular endothelial growth factor
WB: western blot
WES: whole-exome sequencing
WGS: whole-genome sequencing
WNT: wingless-related MMTV integration site
XCI: X-chromosome inactivation
XLH: X-linked hypertrichosis
XLMR: X-linked mental retardation

ACKNOWLEDGEMENTS

Throughout my life, there have been many people who have inspired me to pursue this path and whose support and guidance have been instrumental in both my professional and personal development. The friendships I've cultivated over the past five years and the opportunities I have had to mentor others around me have truly enriched my life. I am grateful to all my family members and close friends, past and present mentors, as well as my colleagues at CUMC, from whom I have learned so much.

I would like to express my deepest gratitude to Dr. Angela Christiano, who enthusiastically accepted me into her lab on June 1st 2011, and has ever since been an incredible mentor and source of support and encouragement. Her confidence and faith in me has significantly impacted my growth as a scientist, and I value the independence she has afforded me to investigate interesting and challenging questions as I pursued my studies in her lab. Throughout the years, Angela has provided me with an abundance of opportunities to develop a wide range of valuable and transferrable skills, from writing manuscripts, book chapters, grants and fellowships, to presenting my research at national and international conferences. I have grown tremendously in the time I have worked in her lab and am greatly indebted to her and all she has done to promote my confidence as a researcher and geneticist, as well as my overall professional development.

During my time at Columbia, I have received an incredible amount of support and encouragement from my committee members, Drs. Virginia Papaioannou and Timothy

Bestor, both of whom served on my qualifying exam and TRAC committees. I am greatly indebted to Ginny and Tim for generously imparting their wisdom and knowledge and providing me with guidance and close mentorship over the past five years. I thank Ginny for serving as my second reader and for always being so supportive of me; her faith in my ability as a scientist has been an invaluable source of encouragement for me. I am truly grateful to Tim for being an instrumental part of my transformation from student to scientist, for graciously welcoming me into his lab both during and following my rotation, and for his continued mentorship, encouragement and friendship throughout the years. I am especially grateful for the wonderful opportunity he provided me with to TA his Advanced Eukaryotic Molecular Genetics course.

I would like to thank Dr. Dorothy Warburton for serving on my qualifying committee and providing helpful feedback on my project, and would like to thank her as well as Dr. Brynn Levy for collaborating with us in the cytogenetic analyses of the hypertrichosis cases in Chapters II and III of my thesis. I also thank Brynn for serving on my thesis defense committee, and extend my gratitude to Dr. David Goldstein for his support and expertise while serving on my thesis defense committee as well.

I am grateful to all the professors who contributed their time and effort to teach our genetics courses, those faculty members who have always encouraged me, and for every member of this department who has truly made my experience here memorable. I will forever cherish the exciting memories of our departmental retreats and how they always refreshed my mind, so I thank every member of the Genetics department who played a role in organizing our retreats. I am grateful to the administration of Genetics and Development who work so hard to keep our department running smoothly. I would

especially like to thank Stacy Warren, Celia Morales, and Jessica Sama for their support and assistance throughout the years with my F31 fellowship, and for always being pleasant and enjoyable to interact with.

I would like to thank the administrators Chantal Manchester and Arthur Uhimov of the Basic Research Science Research Group (BSRG) in Dermatology for all their hard work and for always being so helpful and resourceful. It has truly been a pleasure working with them over the past four years. I am also grateful to members of the Christiano laboratory for engaging me in insightful discussions, fostering my intellectual creativity, and to those who have been supportive and generous with their time in teaching me new methods or answering my questions. I thank Drs. Noriko Arao and Satoru Shinkuma for their assistance and for encouraging me to be hard working and persistent by leading by example. I am especially grateful to our wonderful technician, Ming Zhang, who works day and night to keep the lab running smoothly. His exceptional work and technical assistance has been instrumental in moving my projects along, but most importantly, his amiable personality and sense of humor has truly made working in the lab an enjoyable (and entertaining) experience.

I would like to express my gratitude to our collaborators and all who have provided me with assistance, feedback, troubleshooting advice and reagents during my time in the lab. I am especially grateful to our dermatologist and collaborator from Mexico, Dr. Julio C. Salas-Alanis, who has graciously provided us with patient materials for many of our hypertrichosis studies. It has truly been a wonderful experience working and interacting with him throughout the years, and his continuous support and friendship is invaluable to me. I would also like to thank our collaborator, Dr. Kwame

Anyane-Yeboa, who provided us with patient materials for the autosomal recessive CGHT case, and I am grateful to him for inviting me to present my research at the CUMC Genomic Medical Case Conference last year. I extend my gratitude to Drs. Claudia Dall'Armi and Gil Di Paolo for collaborating with us on the ABCA5 project in the autophagy and lysosome CGHT studies, and for their insightful feedback on the manuscript. I would very much like to thank my colleagues and dearest friends of the Bestor lab, Mathieu Boulard and Olya Yarychkivska—two amazing scientists who have been incredibly helpful to me long after I rotated in the laboratory. I thank Mathieu for his invaluable mentorship during my rotation and for sharing his time and resources to foster my development as a scientist. I learned so many valuable skills and techniques from him that became essential for the design and implementation of my thesis research.

Throughout the past five years, I have met some really amazing people and developed long-lasting friendships with my peers and colleagues. From the very start, my impression of Columbia and the Genetics and Development program has always been so positive, greatly influenced by the many brilliant-minded, sociable and vivacious graduate students who took me under their wings and continued to provide support and advice in the years following. I express my deepest gratitude to Olya Yarychkivska, my dearest friend and partner-in-crime, who has always been so tremendously supportive and a source of encouragement and happiness in my life. You have made my experience at Columbia truly amazing on so many levels, and I will forever cherish our many laughs and amusing moments that electrified the 14th floor of Hammer! To my dearest Elena—my partner-in-crime and classmate—you always have been such an

amazing, dependable friend and a source of support for me over the years. I can't imagine what my time here would have been like without you, and I thank you for always inspiring me, making me laugh and most importantly, keeping me sane! I am so grateful to my other wonderful classmates Casandra Panea, Rory Coleman, Alex Paul, and Michael Smith for truly enriching my experience here, and from whom I have learned so much. I will always cherish the memories, laughs, adventures, the many scintillating scientific discussions, and of course, our brilliantly written skit!

I would also like to thank my dearest friends both within and outside of Columbia, Annina DeLeo, Kerri Carroll, Brittany Maiolo, Antonia Broughton, Megan Furniss, Lisa Kennedy, Yanne Doucet, Maria Salle, Danny Concepcion, Tiffany Zee, Joyce Hui-Yuen, and many, many others for always being a source of positivity and encouragement in my life when I most needed it. I am truly grateful to Joseph Soliman for inspiring and uplifting me with faith and hope, reminding me to strive for the best in myself and share my greatest qualities with the world around me. You have really been a bright soul in my life and I am grateful for the joy you have brought into my world.

I am thankful for all my past and present mentors who inspired me to pursue this path. My high school biology teacher, Mrs. Lynch, played a major role in my decision to pursue a major in biology during college and to this day, remains a dear friend of mine. My college genetics professor and undergraduate research mentor, Dr. Laura Vallier inspired me with by her passion for her work and dedication to mentoring. She taught me everything I knew about genetics and was my primary advocate when applying to graduate school and this program. Her support and friendship has had a lasting influence on my life and how I mentor students.

The most important support of all has come from my family. From the start, they have encouraged me to pursue my every endeavor and ambition, regardless of how great the leap. I am truly blessed to have the greatest parents in the world who continuously offer their most thoughtful, valuable advice, and wise words of wisdom that have shaped and molded my character throughout the years. Though not scientists themselves, my parents have always expressed an interest in my work (and tried their very best to understand it!), and have enthusiastically adhered to my side through my entire experience at Columbia—from the most exciting moments to the greatest struggles and challenges. I am so very blessed to have the two most amazing siblings, John and Jackie, whose lively personalities, kind-hearted and generous nature (and quirky sense of humor!) have always been such a great source of joy in my life. I would not be the person I am today without their love and support.

DEDICATION

To my loving parents who have always fostered my creativity and encouraged me to strive for the best in myself, pursue my most ambitious dreams, and most importantly, share my gifts with the world around me. I am truly blessed and eternally grateful for their unconditional love and support in my every endeavor.

~*~

To all who inspire others with their enthusiasm and passion

to dream, believe, achieve, and succeed.

Chapter I

General Introduction

Chapter I. General Introduction

Chapter I.1. Patterns of Inheritance

In humans, genetic disorders can arise from a variety of mechanisms and mutational events in the genome, ranging from changes in single nucleotides to structural rearrangements of chromosomes. Beyond mutations, epigenetic or heritable changes in gene expression and phenotype can also play a key role in disease pathogenesis. Thus, we can use inherited conditions as models to identify novel genes and mechanisms controlling a particular trait of interest (i.e. hair growth). However, the approaches used to identify pathogenic mutations in a given disorder depend on the inheritance pattern and disease prevalence in the population (Figure 1). To date, there are approximately 7,000 rare monogenic diseases reported in Online Mendelian Inheritance of Man (OMIM) (defined as affecting fewer than 1 in 200,000 individuals in the United States or fewer than 1 in 2,000 in Europe), yet the molecular underpinnings of nearly half of these diseases has been identified. While the genetic architecture of some Mendelian diseases may have been more readily detectable, the pathogenic variants underlying several rare genetic disorders have been difficult to identify using the traditional approaches, due to limited availability of affected families, decreased reproductive fitness, and locus heterogeneity. Similarly, complex (polygenic) disorders have not been amenable to the conventional methods and approaches. Importantly, the completion of the Human Genome Project in 2003 greatly facilitated the development of advanced technologies for genome sequencing and mutational discovery in both monogenic and polygenic conditions.

Based on the inheritance pattern and disease prevalence in the population, human genetic disorders can primarily be classified as: 1. Rare, sporadic, 2. Rare, familial Mendelian, and 3. Complex, polygenic (non-Mendelian). In the rare, sporadic Mendelian traits, the genetic architecture is usually characterized by large genomic rearrangements or copy number variants (CNVs) that are private, or restricted to the proband and immediate relatives (GILISSEN *et al.* 2012), whereas in rare, familial conditions, the genetic architecture is composed of intragenic (i.e. exonic, splice site) mutations that are not private but rather, tend to be present at a very low frequency (~0.1-0.5%). Both of these diseases are characterized by rare variants contributing large effect sizes. In contrast, in common, polygenic conditions, many common and low frequency variants (1-5%) contribute small effects and interact with the environment to produce the phenotype (CIRULLI AND GOLDSTEIN 2010). The detailed approaches and methods to identify pathogenic mutations are discussed in Chapter I.1.3.

I have used several methods and approaches to interrogate the genetic and molecular basis of three inherited conditions affecting hair growth, and obtain novel insight into mechanisms and pathways required for normal hair growth. The three approaches to study rare, Mendelian (sporadic, familial) and non-Mendelian disorders comprise the basis of my thesis and are discussed in Chapters II-IV.

1.A Mendelian (Monogenic) Inheritance Patterns

A Mendelian-inherited trait is one controlled by a single gene or locus and transmitted to offspring in a consistent pattern and frequency based on the laws of segregation and

independent assortment of chromosomes.

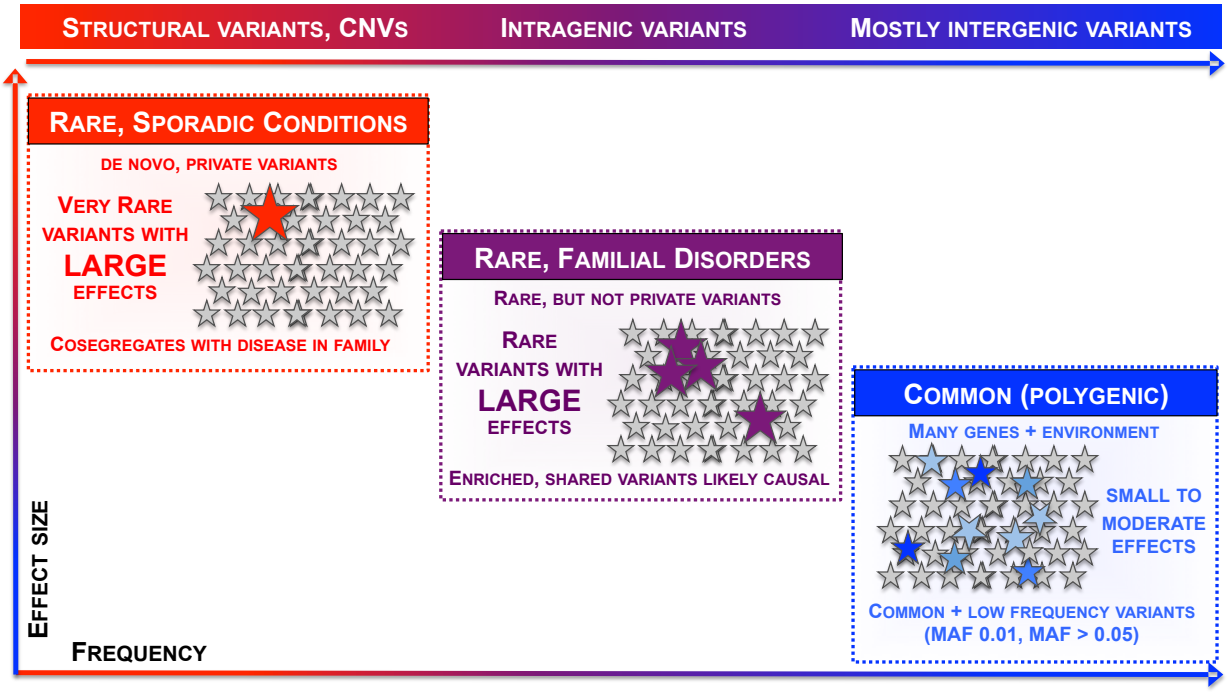


Figure 1. The genetic architecture of Mendelian and non-Mendelian disorders. Rare, sporadic conditions (red) are characterized by *de novo*, private variants (structural variants, CNVs) with large effect sizes. Rare, familial disorders (purple) are most often composed of rare, intragenic variants with large effect sizes, and non-Mendelian, polygenic conditions (blue) are characterized by many common and low-frequency variants with small effects, plus environmental interactions.

The genetic variants underlying monogenic (single-gene) traits can be located either on autosomes, producing an autosomal inheritance pattern, or on the X chromosome producing an X-linked inheritance pattern. When a mutation in one of two alleles is sufficient to produce a phenotype, the trait is considered dominant, whereas a condition is recessive when both alleles of a gene must be altered in order to produce a phenotype. Thus, a Mendelian-inherited trait can be classified into four groups: 1. Autosomal dominant, 2. Autosomal recessive, 3. X-linked recessive, and 4. X-linked dominant.

An autosomal dominant trait can be transmitted from either parent to the offspring with the same frequency, but not from an unaffected individual. The mechanisms whereby dominant mutations can give rise to a phenotype are haploinsufficiency, where one gene product is not enough to sustain function, and a dominant negative effect, where the mutated protein interacts with and prevents the wild-type protein from functioning. In a pedigree of an autosomal recessive disease, affected individuals have clinically unaffected parents, and there is no sex bias or vertical transmission. The mechanisms whereby recessive mutations can give rise to a phenotype include familial consanguinity that leads to homozygosity (identical by descent) for the mutant allele, and compound heterozygosity for two distinct mutations in the same gene. Both phenomena have been observed in several skin and hair disorders, yet familial consanguinity is the more common mechanism underlying rare traits aggregated in certain geographic populations. In Chapter II, I investigated the genetic basis of an autosomal recessive condition of excessive hair overgrowth in a consanguineous family.

In X-linked dominant traits, affected individuals have at least one affected parent, the disorder can occur in both males and females, and there is no male-to-male transmission. Some X-linked dominant diseases are lethal in males during early intrauterine development but occasional male patients can be explained by functional or genomic mosaicism arising from an extra X chromosome or a *de novo* half-chromatid/postzygotic mutation, respectively. Females with X-linked conditions are considered hetero- or hemizygous (for dominant and recessive traits, respectively) and generally have a milder or mosaic phenotype. Lyonization, or random epigenetic

inactivation of one of the two X chromosomes in each female cell produces this mosaic pattern, which can be directly visualized in skin diseases. This phenomenon has been observed in females with focal dermal hypoplasia (Goltz syndrome), characterized by mutations in the *PORCN* gene (GOLTZ 1992) and incontinentia pigmenti, caused by mutations in the *NEMO* gene (SMAHI *et al.* 2000). In Chapter II, I examined the genetic basis of an X-linked condition of excessive hair overgrowth using the approaches and methods discussed in the following section (Chapter I.2.A).

1.B Exceptions to Basic Mendelian Inheritance

Mendelian diseases are often caused by mutations in a single gene or locus, however, there are several mechanisms by which this gene can modify/be modified by others and by epigenetic/environmental factors, creating a more complex and variable phenotype. As such, there are several monogenic conditions that appear to defy the basic Mendelian patterns of inheritance due to variation in gene expression, the location of the mutant gene, as well as other factors, and thus have major impacts on risk assessment for disease transmission and genetic counseling. These deviations from the basic Mendelian patterns are considered exceptions, and include incomplete or reduced penetrance and late-onset diseases, variable expression, X inactivation, *de novo* mutations, genomic and gonadal, revertant mosaicism, inbreeding, quasidominant inheritance, uniparental disomy, mitochondrial inheritance, epigenetics and imprinting. A thorough explanation of these exceptions to basic Mendelian patterns and examples in monogenic disease is described in detail elsewhere (DESTEFANO AND CHRISTIANO *(In press)*) (DESTEFANO AND CHRISTIANO 2014).

Throughout the work described in my thesis, the approaches and methods I have utilized to identify the underlying pathogenic mutations in conditions affecting human hair growth have been guided by the inheritance pattern and knowledge of any exception(s) to basic Mendelian patterns. Chapters II and III discuss *de novo* mutations and consanguinity in two inherited forms of excessive hair growth.

1.C Chromosomal Anomalies, Structural Variants, and Position Effects

In many human disorders, the underlying genetic defect is not contained within a single gene but rather involving a large genomic region or even entire chromosomes. Historically, researchers detected numerical and structural chromosomal anomalies using *karyotypes*, which are chromosome "profiles" that allow for visualization of chromosome banding patterns via a chemical stain that detects densely packed DNA (heterochromatin) and loosely packed DNA (euchromatin) in metaphase chromosomes. The current methods of detection for chromosomal abnormalities are discussed in Chapter I.2.A.

Chromosomal anomalies are defined as changes in the number and/or structure of genomic DNA fragments. Chromosomal numerical abnormalities can involve the whole genome (polyploidy) or just one pair of homologues (aneuploidy). Chromosomal structural anomalies involve chromosome breakage and create a net gain/loss (unbalanced) or no change (balanced) in genetic information, where the latter can include translocations and inversions. In a reciprocal translocation, two breaks in two non-homologous chromosomes occur and the chromosomal segments distal to the breaks are exchanged, and the total number of chromosomes remains the same. If the translocation breakpoint affects a gene/regulatory sequence or places it under the

control of novel regulatory elements (i.e. in the case of the *MYC* gene in Burkitt's lymphoma) (BERNHEIM 2010), it can create a position effect (discussed below) on nearby genes and cause an observable phenotype. In contrast to translocations, inversions are balanced rearrangements in which two breaks occur in the same chromosome and the intervening segment is inverted before the chromosome is reconstituted. In a *pericentric inversion*, the inverted segment includes the centromere, whereas in a *paracentric inversion*, the two breaks occur in the same chromosome arm, without involvement of the centromere.

A classic example of chromosomal anomalies associated with human genetic disease is hypohidrotic ectodermal dysplasia (HED), characterized by a decreased ability to sweat, where patients have been reported to possess rearranged X chromosomes, including X;autosome translocations or genomic deletions involving the *EDA* gene (THOMAS *et al.* 1993). Similarly, CNVs on chromosome 4q26-q27 have been reported in patients with Cantú Syndrome, a rare disorder of congenital hypertrichosis, cardiomegaly, macrosomia, and osteochondrodysplasia (KURBAN *et al.* 2011). Cytogenetic abnormalities have also been reported on chromosome 8q22 in several cases of Ambras syndrome (BAUMEISTER *et al.* 1993) (BALDUCCI *et al.* 1998) (TADIN *et al.* 2001; FANTAUZZO *et al.* 2008), a form of congenital universal hypertrichosis, which the Christiano lab has previously studied (discussed in Chapter I.5 and I.6)

Interestingly, changes in copy number and/or location of chromosomal segments can disrupt the expression of the neighboring genes, leading to position effects. A position effect is defined as a change in gene expression as a result of altering the location of the gene relative to its chromosomal surroundings (KLEINJAN AND VAN

HEYNINGEN 1998). Several genetic mechanisms, including chromosomal rearrangements, insertions, deletions, and copy number variants, can lead to the separation of the gene from a tissue- or temporal-specific modifier element (KLEINJAN AND VAN HEYNINGEN 1998). This definition of position effect is distinct from the classical phenomenon of position effect variegation, first reported in *Drosophila* at the *white*$^{+}$ locus, which describes the heritably stable inhibition of gene expression as a result of juxtaposition with heterochromatin (WEILER AND WAKIMOTO 1995).

Paradoxically, a position effect on a gene may cause a different phenotype from that produced by loss of function mutations in the coding region of the gene. This, in conjunction with the fact that most diseases are relatively rare make it more difficult to identify and directly implicate position effects in human genetic disease. Despite these obstacles, position effects on single genes have been reported as a result of large chromosomal rearrangements or copy number variants in several human genetic diseases. Some examples include *SHH* in Holoprosencephaly (a cephalic disorder in which the forebrain fails to develop into two hemispheres), *SHH* in Preaxial polydactyl II (a congenital physical anomaly resulting in extra digits), and *SOX9* in Campomelic dysplasia (a genetic disorder of the femur and tibia) (ROESSLER *et al.* 1997; LETTICE *et al.* 2003; VELAGALETI *et al.* 2003; FERNANDEZ *et al.* 2005; LEIPOLDT *et al.* 2007).

The Christiano lab has previously reported position effects on genes regulating hair follicle development, which will be discussed in Chapter I.5. Throughout the work described in Chapters II-IV, chromosomal anomalies, structural variants, and position effects are discussed in the context of genetic mechanisms regulating human hair growth, which is the subject of my thesis.

1.D Polygenic Disorders

Human genetic disorders that are caused by mutations in a single gene are quite rare within the general population, representing only a small proportion of genetic diseases. In contrast, the majority of inherited conditions are more common, genetically complex, and do not follow the basic Mendelian inheritance patterns despite their high prevalence within families; these are referred to as polygenic disorders or multifactorial traits. In addition to multiple genetic factors underlying the pathology, genetic interactions with the environment can significantly impact the disease prevalence amongst related individuals and in the general population, as well as phenotypic severity (i.e. expressivity). Thus, the risks for multifactorial traits have been estimated based largely on epidemiologic data. It is currently thought that the genetic basis of complex human disorders is composed of many common variants with small to moderate effects, rare variants with moderate to large effects, structural variations (i.e. CNVs), genetic interactions, and gene-environment interactions (MANOLIO *et al.* 2009) (MACARTHUR *et al.* 2014). Although the classical genetic approach (discussed in Chapter 1.2.A) has been widely successful in identifying the pathogenic genes underlying diseases inherited in a Mendelian fashion, these methods have not been as effective in dissecting the genetic architecture of polygenic diseases. However, powerful tools have been developed over the past decade that circumvent the limitations of preexisting methods and have aided in uncovering the genetic complexity of many of these diseases (discussed in Chapter I.2.B).

In common, complex diseases, the degree of inheritance was usually inferred by studying the trait in monozygotic (identical) and dizygotic (fraternal) twins, where a high

discordance rate between monozygotic twins strongly suggests environmental and/or epigenetic factors that contribute to the variability in inheritance and expressivity of the trait (ZWIJNENBURG *et al.* 2010). For example, in alopecia areata, an autoimmune disease of hair loss, the concordance rate among monozygotic twins is 55%, whereas that in dizygotic twins is 0%, strongly suggesting a genetic basis (RODRIGUEZ *et al.* 2010). In addition to twin studies, genetic association studies and parametric linkage analyses using epidemiological data identified regions of the genome significantly associated with the trait (PETUKHOVA AND CHRISTIANO 2013). Collectively, twin studies, family-based linkage analyses, as well as observed heritability among relatives were used to provide evidence for a genetic basis in common, complex diseases.

Nearly a decade ago, the Genome-Wide Association Study (GWAS) became the gold standard technique for identifying common SNPs shared among unrelated individuals with a given complex trait. This method assays several thousand to more than a million SNPs in the genome, requiring a large sample size to produce a high resolution and power for detection of differences in allele frequencies between control and affected individuals. Several GWAS studies have been performed on at least eleven polygenic skin and hair related diseases within the last six years (Petukhova et al. 2010; Sarig et al. 2012; Zhang 2012; Luo et al. 2013), confirming previous findings from parametric linkage analysis and genome-wide scans as well as identifying new risk variants. These methods will be further discussed in the following section of Chapter I.2.B.

Chapter I.2. Approaches to Identifying Pathogenic Mutations

2.A Approaches and Methods to Identify Disease-Causing Mutations in Rare Mendelian Diseases

One overarching goal in the study of inherited diseases is to find the causative gene(s) responsible for the phenotype manifested in a given disorder and define its function. The classical genetic approaches used to map, sequence, and characterize mutations (discussed below) have proven to be the most fruitful in identifying hundreds of genetic mutations in dozens of monogenic Mendelian diseases (Figure 2). However, for rare Mendelian conditions that are sporadic and in complex traits (discussed in Chapter I.2.B), the underlying genetic lesions were difficult to identify using the conventional methods (Figure 2). Importantly, next-generation sequencing technologies, in conjunction with CNV analysis (discussed below) have been instrumental in identifying *de novo* chromosomal structural defects in single affected individuals as well as within families.

In rare, Mendelian extreme traits that are sporadic, the genetic architecture is usually characterized by large genomic rearrangements or copy number variants (CNVs) that are private, or restricted to the proband and immediate relatives, producing large effects (i.e. disrupting multiple genes) (Figure 1). In the cases where linkage and cytogenetic methods to detect chromosomal abnormalities provided an incomplete picture of the genetic architecture, Next-Generation Sequencing (NGS) methods including whole-exome and whole-genome sequencing have been successful in uncovering novel and rare variants (Figure 2). Since the mutational event is sporadic and private, determining causality is limited to performing a cosegregation analysis,

whereby all affected family members are genotyped for the candidate variant(s) to determine whether it fully cosegregates with the disease phenotype.

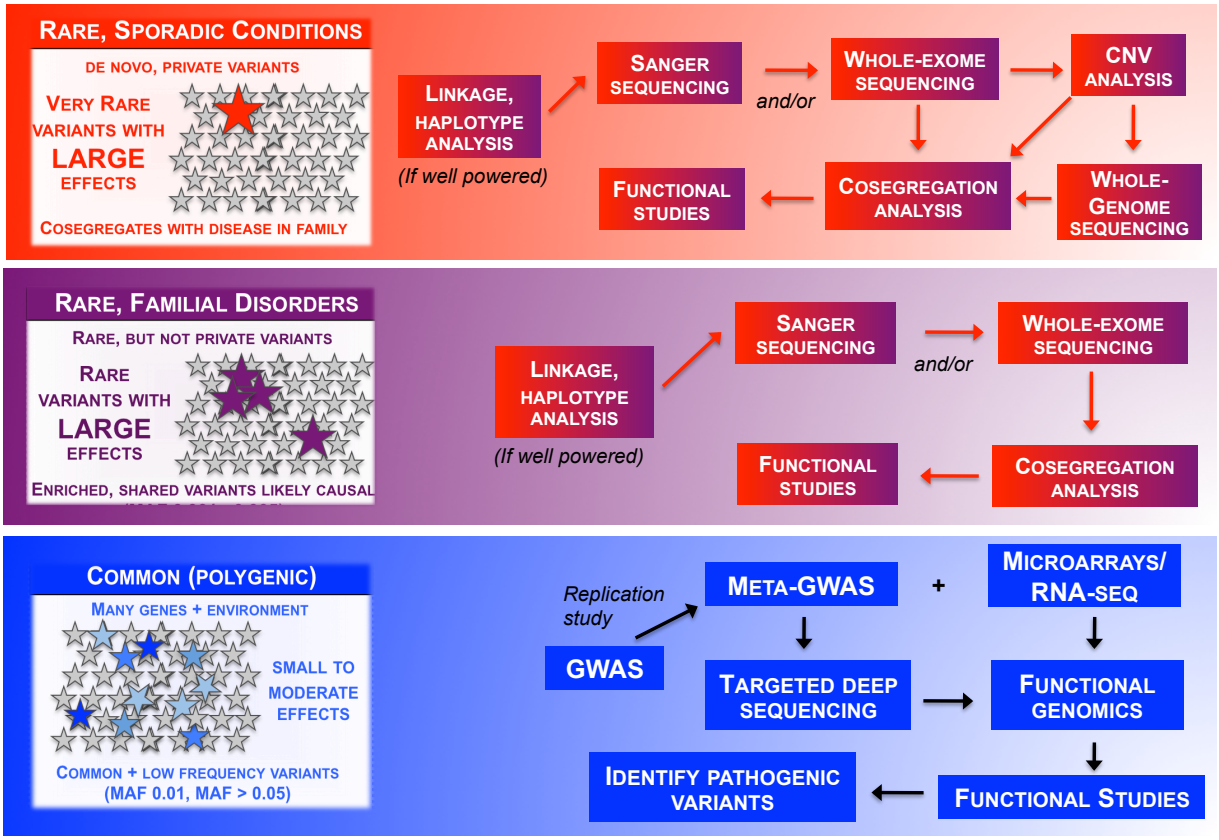


Figure 2. Approaches and Methods to Identify Disease-Causing Mutations. In rare, Mendelian disorders of the skin and hair, the genetic architecture is often composed of rare variants with large effects, detected by linkage, CNV analysis and/or whole-exome and genome sequencing. In contrast, polygenic disorders are composed of common and rare variants that contribute small to moderate effects and can be detected using Genome-Wide Association Studies (GWAS), followed by functional genomics studies.

In the rare, Mendelian traits that are familial, the genetic architecture is composed of intragenic (i.e. exonic, splice site) mutations that are not private but rather, are present at a very low frequency (~0.1-0.5%) in the population, exerting large effects (Figure 1). These variants are detectable by the conventional methods of linkage and Sanger sequencing for gene mapping (discussed below), and family cohorts are used to determine cosegregation of the variant(s) with the disease phenotype (Figure 2). While these shared variants are likely causal, their identification in independent and unrelated

family cohorts manifesting the same phenotype, in addition to functional studies (i.e. mouse models and/or cell culture based systems) provide the strongest evidence for their causality in disease pathogenesis.

Upon the identification of candidate variants in rare Mendelian disorders using the approaches discussed below, cosegregation analysis using Sanger sequencing will establish whether the variant is present in the DNA of all affected family members, thereby cosegregating with the disease. Once cosegregation is established, downstream functional analyses should be performed to determine the involvement of the variants identified in disease pathogenesis.

Functional and Positional Cloning Strategies in Mendelian Disorders

Two major approaches have been used to isolate the gene responsible for a given phenotype: functional and positional cloning (Figure 3) (YAMANISHI 1996; DESTEFANO AND CHRISTIANO 2014). The suitability of each approach for a given disease depended on what prior biochemical information was known about the defective gene product, as well as whether the inherited disease was sufficiently prevalent to study multiple affected individuals within a kindred. Functional cloning (reverse genetics), or the candidate gene approach, does not require positional information but rather exploits information on the defective pathway, function, structure or tissue underlying the phenotype.

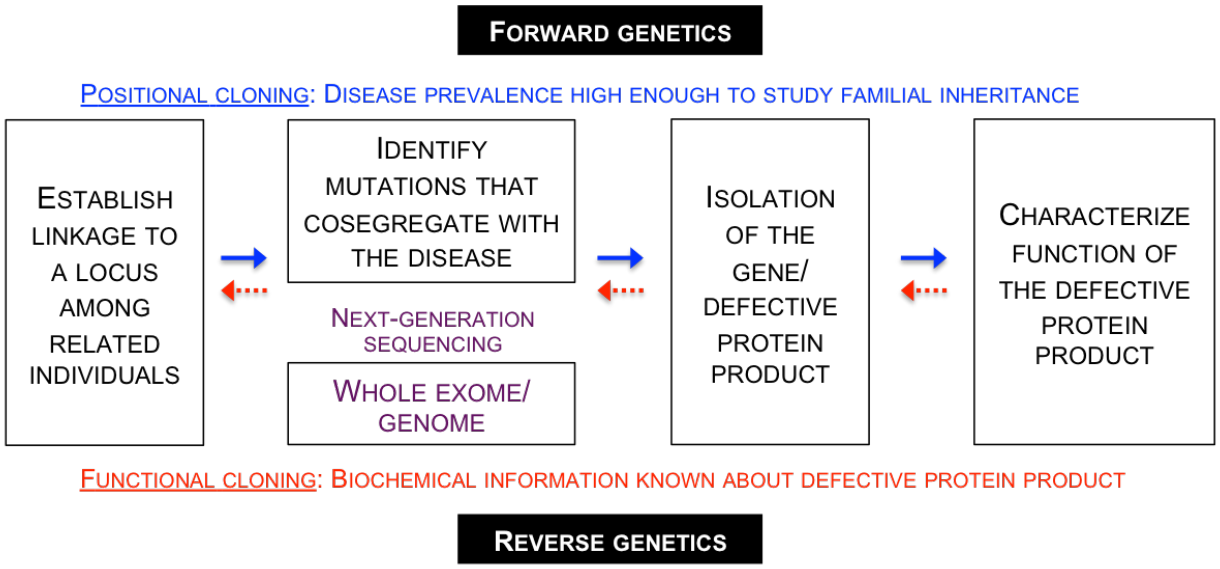


Figure 3. The classical genetic approach of functional and positional cloning to study Mendelian diseases. In positional cloning, or forward genetics, inheritance of a given trait is prevalent enough in related individuals to identify the shared genetic region. In functional cloning, or reverse genetics, information regarding the functional properties of the protein product is exploited to then isolate and map the gene. (Adapted from DeStefano and Christiano, 2014)

To select and clone a candidate gene, knowledge of the defective protein product/biochemical pathway and homology to mutated genes responsible for closely related phenotypes are important criteria. Animal models have also proved useful resources in elucidating the candidate gene, an approach that relies on the existence of a syntenic chromosomal region, where the spatial arrangement of encompassing genes is evolutionarily conserved. Once the genetic interval in humans is determined, the syntenic interval in mice can be analyzed to identify additional candidate genes. If a similar phenotype in the mouse has been mapped to this region and the mutated gene is known, this becomes an excellent candidate for the human disease counterpart, and vice versa. This comparative genomic approach was used to identify mutations in the *Hairless* gene underlying atrichia with papular lesions (AHMAD *et al.* 1998) and

mutations in the *DSG4* gene in patients with localized autosomal recessive hypotrichosis (KLJUIC *et al.* 2003), where *Hairless* and *DSG4* mutations in mouse models phenocopied the human diseases. Although functional cloning has been useful for elucidating the pathological and genetic basis of several monogenic diseases, its application is limited to the few disorders that feature an obvious biochemical defect.

In contrast, positional cloning (forward genetics) relies only on the inheritance of the disease among related individuals to identify the common, shared region that completely cosegregates with the disease phenotype (gene mapping) (Figure 3). Once the candidate region has been narrowed, the gene can be isolated and cloned, and further studies can be performed to determine its function. One of the most widely used methods to accomplish this is linkage analysis, a tool that enables the mapping of a gene or locus by analyzing its segregation with genetic markers (microsatellite markers, SNPs) of known chromosomal position. These methods will be discussed in greater detail in the following section.

The approach that has been most fruitful in identifying candidate disease genes is a combination of both functional and positional information. In the positional candidate approach, once the chromosomal location of the disease locus is known, functional information is utilized to directly analyze the candidate genes contained in that particular chromosomal region. Among many human genetic disorders, the positional candidate approach was successfully employed to investigate the genetic basis of hypertrichosis. The Christiano lab recently reported homozygous recessive mutations in the *FGF5* gene, a known regulator of hair length, in large consanguineous families from Pakistan with trichomegaly (long eyelashes) and hypertrichosis (HIGGINS *et al.* 2014). Linkage

and homozygosity mapping (discussed below) were conducted using SNP-genotyping to identify shared regions of the genome for which affected individuals were homozygous, and whole-exome sequencing identified intragenic mutations in *FGF5,* defining the candidate region (HIGGINS *et al.* 2014).

Importantly, the human genome sequence as a resource has made the positional candidate approach the most powerful disease-gene cloning strategy because it provided a detailed map of polymorphic markers and positional information on characterized (known) as well as predicted genes. Publically available Internet resources include the National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov), Ensembl Genome Browser (www.ensembl.org/), and the UCSC Genome Browser (http://genome.ucsc.edu/index.html).

Linkage and Haplotype Analysis

The power of linkage is based on the genetic principles of segregation and independent assortment, such that two traits located on separate chromosomes (or very far apart on the same chromosome) segregate independently, and thus have a recombination fraction of 50%. The closer the traits are, the less frequent recombinants are found; a recombination fraction (RF) of 0 implies the traits are tightly linked. In pedigree analysis, LOD (logarithm of the odds) scores are generated to calculate the likelihood that two loci are linked (MORTON 1955); positive scores indicate linkage (≥ 3.0) and negative scores indicate that linkage is less likely (≤ -2.0). To conduct linkage analysis, a thorough definition of the phenotype, availability of pedigrees with sufficient power for linkage, as well as map of polymorphic SNP markers are required.

Once the linkage interval has been defined, an additional step of narrowing the locus is often required since these intervals often span tens-to-hundreds of kilobases encompassing hundreds of genes. In some cases, information on a candidate gene and/or chromosomal abnormalities associated with the disorder is available and can aid in refining the interval. In the absence of this information, the number of individuals and recombination events/markers analyzed should be increased to define the shared haplotype. Importantly, recombination events can 'break' the shared haplotype in different members/generations of the pedigree, so identifying an individual in whom recombination has occurred *inside* the disease-associated haplotype (the key recombination event), further reduces the region where the disease gene is presumed to be located.

The linkage interval that cannot be further reduced is considered the candidate region that cosegregates with the disease phenotype and harbors the disease gene. Sequencing and genotyping-based methods can then be used in conjunction with haplotype analysis to analyze the candidate regions and identify genetic variants. In the case of autosomal recessive traits driven by rare alleles with large effects, regions of homozygosity can be mapped using SNPs as markers for homozygosity, or autozygosity mapping. Moreover, overlaying identical-by-descent regions of the genome with linkage peaks is a robust approach and has been fruitful in the identification of candidate genes for several conditions (GILISSEN *et al.* 2012) (PETUKHOVA *et al.* 2009).

Identifying Structural Variants by CNV Analysis

The genetic architecture of rare, sporadic conditions is usually characterized by large genomic rearrangements or copy number variants (CNVs). Historically, researchers

have used karyotyping to detect numerical and structural chromosomal anomalies, however, fluorescence *in situ* hybridization (FISH) has gradually become a preferable choice, as it detects DNA sequences as short as 1 Kb (MILLER 2001) and allows for the resolution of very small changes in the chromosomal constitution that are not visible by the use of conventional banding techniques. To detect multiple chromosomal regions simultaneously, multicolor FISH (mFISH) allows for combinatorial labeling using fluorophore-conjugated probe sets, and for whole-genome analysis, Comparative Genomic Hybridization (CGH), enables the identification of all CNVs in a single hybridization procedure, and maps these changes onto normal chromosomes (PINKEL AND ALBERTSON 2005). While this method has an improved resolution (5-10 Mb) compare to chromosomal banding, it still requires metaphase chromosomal preparations and is unable to identify balanced chromosomal rearrangements. More recently, advances in microarray technology have enabled the use of DNA microarray-based CGH (aCGH) for detection of copy number variations (CNVs) at a considerably higher level of resolution (~100 Kb) than chromosome-based methods, making it more suitable for diagnostic applications (GAO *et al.* 2012).

Currently, SNP oligonucleotide microarray analysis (SOMA) yields the highest resolution (5.8 Kb), combining SNP data with the detection of CNVs (TORGYEKES *et al.* 2011). SOMA is widely used for clinical cytogenetic analysis because both visible and submicroscopic cytogenetic imbalances are detectable, which has allowed for the detection of cryptic insertions/deletions on apparently balanced chromosomes. Additionally, genotype information obtained from mapping SNP positions to the genome browser is available for further DNA-based studies including uniparental disomy and

zygosity. Moreover, this method can identify all previously characterized cytogenetic aberrations (TORGYEKES *et al.* 2011).

In the work described in Chapters II and III, these cytogenetic methods in conjunction with next-generation sequencing methods (discussed below), were used to identify the underlying genetic lesions in human conditions characterized by excessive hair growth.

Whole-Exome and Whole-Genome sequencing

When studying Mendelian traits, including those not amenable to conventional genetic approaches, it is of critical value to identify the rare variants that cosegregate with the disease phenotype. In such cases where linkage is not applicable or was not informative, whole-genome and whole-exome sequencing methods have proven very useful in identifying pathogenic variants that were otherwise missed. Next-generation sequencing in probands of large families with linkage evidence has also proven a useful strategy to narrow the list of candidate genes and identify pathogenic mutations, similar to the approach of overlaying information on identical-by-descent with linkage peaks described in the previous section.

The coding fraction, or exome, comprises only 1-2% of the genome, however, its contribution to biological function is profound, evidenced by the large number of genetic diseases caused by mutations in single genes (GILISSEN *et al.* 2012). In Mendelian traits for which there is known linkage information and/or a candidate gene, whole-exome sequencing can be used to identify variants in coding regions, an approach that has been highly fruitful in identifying rare variants with large effects (CIRULLI AND GOLDSTEIN 2010; LAI-CHEONG AND MCGRATH 2010). For both recessive and dominant traits, whole-

exome sequencing followed by bioinformatics analysis has recently emerged as an efficient, more cost-effective approach to identifying mutations compared to whole-genome sequencing. This method is especially useful for studying monogenic conditions caused by intragenic (exonic, intronic, etc) mutations that produce large phenotypic effects.

In monogenic skin and hair disorders, massive parallel whole-exome sequencing has successfully identified many novel genes in both dominant and recessive conditions, including hypotrichosis simplex (*RPL21* gene, autosomal dominant), terminal osseous dysplasia (*FLNA* gene, X-linked dominant), Adams-Oliver syndrome (*ARHGAP31* gene, autosomal dominant), Clericuzio-type piokiloderma with neutropenia (*c16orf57* gene, autosomal recessive), and Sensenbrenner syndrome (*FLNA* gene, X-linked dominant) (LAI-CHEONG AND MCGRATH 2010). Importantly, whole-exome sequencing in conjunction with homozygosity mapping has been shown to be key for studying autosomal recessive conditions for which there is known consanguinity and the potential involvement of multiple genes. This strategy was recently demonstrated in a case of oculocutaneous albinism and neutropenia, an atypical disease phenotype comprised of two separate disorders, where pathogenic homozygous recessive mutations were identified in both the *SLC45A2* and *G6PC3* genes that explained the oculocutaneous albinism and neutropenia phenotypes, respectively (CULLINANE *et al.* 2011). If the candidate approach alone had been used to screen for all the known genes associated with either of these conditions, the genotype-phenotype correlation would have been incorrect and the patient may have been misclassified. More recently, whole-exome sequencing and homozygosity mapping identified a novel homozygous

missense mutation in the *CHST8* gene (encoding the GalNAc4-ST1 transferase) in a large consanguineous family with generalized peeling skin syndrome (PSS) (CABRAL *et al.* 2012). *CHST8* is normally expressed in the epidermis and functional *in vitro* studies revealed that this mutation leads to a significant reduction in protein levels as well as total sulfated glycosaminoglycan levels in the cell, suggesting a complete loss-of-function (CABRAL *et al.* 2012).

In addition to inherited monogenic conditions, whole-exome sequencing has also been successfully used for the identification of *de novo* mutations that are not inherited, but rather acquired in somatic cells. In individuals with Proteus syndrome, postzygotic, somatic mutations were identified in the *AKT1* oncogene, where sequencing reads from the whole-exome study revealed the presence of mosaicism for the mutation with levels varying from 1-50% in affected vs. normal tissues from patients (LINDHURST *et al.* 2011). Importantly, this was the first application of whole-exome sequencing to identifying *de novo* somatic mutations underlying a human genetic disorder.

As a tool to systematically identify genomic aberrations in the context of tumorigenesis, whole-exome sequencing revealed a novel *de novo* mutation in the *GRIN2A* gene (encoding a glutamate receptor) that was frequently observed in melanoma patient samples, and downstream bioinformatics analysis identified the glutamate signaling pathway as significantly enriched in melanoma (WEI *et al.* 2011). Moreover, this approach has successfully identified novel mutations in genes for chronic lymphocytic leukemia, B cell lymphoma, as well as carcinomas of the lung, brain, kidney, and pancreas (LOHR *et al.* 2012) (PUGH *et al.* 2012) (XU *et al.* 2012) (LIU *et al.* 2013) (QUESADA *et al.* 2012). Concurrent with the identification of such pathogenic

variants was the development of analytical tools designed to systematically decode and integrate this information with copy number variation and loss-of-heterozygosity, enabling a robust survey of the genetic landscape of various pathologies (PABINGER *et al.* 2014). Importantly, the availability of these new sequencing technologies and analytic tools has facilitated the discovery of novel, deleterious mutations associated with cancers and other diseases, and in many cases, provided a comprehensive catalog of clinically relevant genes and pathways involved in disease pathogenesis.

Whole-genome sequencing as a tool for disease gene discovery has been used to elucidate the underlying pathogenic variants in human genetic diseases, particularly those with an extreme phenotype and a rare variant enriched in the affected population (CIRULLI AND GOLDSTEIN 2010). This method is especially useful for cases in which these data are not available or uninformative, but used in conjunction with linkage or haplotype analysis, can be helpful in focusing the search for the pathogenic variant. For example, in the rare, dominant Mendelian disease, metachondromatosis, whole-genome sequencing was performed on a patient with partial linkage data, and identified a deletion in the *PTPN11* gene that cosegregates with the disease phenotype (SOBREIRA *et al.* 2010). Moreover, the clinical relevance of *PTPN11* to disease pathogenesis was further validated through the identification of a second mutation in an independent family (SOBREIRA *et al.* 2010). In the context of genetic diagnosis, whole-genome sequencing has been tested in a patient with a recessive form of Charcot-Marie-Tooth disease, and revealed clinically relevant, compound heterozygous variants in the *SH3TC2* gene (LUPSKI *et al.* 2010). Importantly, next-generation sequencing

methods, in conjunction with classical linkage and haplotype analysis provide a robust and powerful approach to identifying pathogenic mutations in rare, Mendelian disorders.

In the work described in Chapters II and III, the next-generation sequencing methods described above were instrumental in the identification of pathogenic variants in rare human conditions of excessive hair growth whose genetic basis had previously remained elusive for decades.

2.B Approaches to Identifying Disease-Causing Mutations in Common, Polygenic Conditions

The candidate gene approach has been successful in identifying pathogenic mutations in Mendelian diseases, however, the more common, polygenic conditions have not historically lent themselves to the conventional genetic approach. Unlike Mendelian disease, their underlying genetic architecture is composed of many common and low frequency variants (1-5%) that contribute small effects and interact with the environment for the full manifestation of the trait (i.e. hair loss). Whereas in Mendelian conditions, family cohorts are used to identify shared variants that are likely causal, population-based cohorts of affected and unaffected individuals are used to identify susceptibility loci in polygenic diseases whose association can be detected by Genome-Wide Association Studies (GWAS) (Figure 2). While most (if not all) of the mutations identified in monogenic (single-gene) diseases reside within protein-coding genes, recent evidence from GWAS, next-generation sequencing and functional genomic studies (discussed in the following sections) have revealed that the majority of significant associations reside within non-coding, intergenic DNA, thus requiring an integrative,

robust analytic approach to determine their contribution to disease pathogenesis (MAURANO *et al.* 2012; GUSEV *et al.* 2014).

Genome-Wide Association Studies (GWAS) and meta-GWAS

For complex disorders in which the trait does not follow a Mendelian inheritance pattern, Genome-Wide Association Studies (GWAS) have been used to identify single nucleotide variants enriched in cases compared to controls. The goal is to identify susceptibility loci and variants that are significantly associated with the trait compared to the general population, and confer a degree of genetic risk. Until recently, this process was challenging, dependent on candidate-based approaches and limiting in its informative value. However, the development of newer technologies (i.e. SNP genotyping platforms) with improved genome-wide coverage has enabled the unbiased, systematic identification of pathogenic variants. Currently, large-scale and genome-wide technologies utilize microarray-based chips containing hundreds and thousands of oligonucleotide SNP probes. The Illumina OmniExpress GWAS array is applicable for genotyping genome-wide variants implicated in GWAS experiments, and specifically for autoimmune diseases, the Immunochip is commonly used to test associations with known disease-associated variants (MOROZOVA AND MARRA 2008).

The success of a GWA study is highly dependent on the nature of the disease being studied (i.e. the characteristics of the phenotype, variation, sensitivity and specificity in the diagnosis, etc), and extremely large cohorts of individuals are required to identify a statistically significant association between a tag SNP (representative of a haplotype) and the disease. Because of this methodology, the associated SNPs only capture of snapshot of the genetic variation across a region thus, requiring follow-up

studies of targeted deep resequencing (discussed below) in candidate GWAS regions to identify the susceptibility variants in association with the disease. It is important to note that some variants, particularly missense and silent substitutions, can be non-pathogenic (present in the general population as polymorphisms), therefore, they need to be crosschecked with publically available databases (1000 genomes, dbSNP, NCBI, UCSC, etc.) containing information on natural genomic variation.

Following GWAS, the next step is to determine the validity of the associations identified by analyzing an independent cohort (i.e. geographic region, ancestry) in a replication study, and then perform a meta-GWAS analysis. A meta-analysis entails combining the data generated from the initial GWAS with that of the replication study, and using statistical methods (i.e. logistic regression) to calculate the odds ratio and standard error per allele. The goal is to replicate the previous findings as well as identify novel loci associated with the trait that surpass genome-wide significance ($p \leq 10^{-8}$). While a replication study is not always available may not have been performed, the meta-GWAS has become the standard approach to validating the association of a genomic region with the disease prior (or in parallel) to embarking upon functional genomics or deep sequencing studies.

The application of GWAS and meta-GWAS to study common diseases is based on the premise that common allelic variants are present in greater than ~5% of the general population, yet the proportion of heritability accounted for by the identified variants is relatively small (~20%) (MANOLIO *et al.* 2009). The “missing heritability” is said to be explained by a large number of variants with small effects, rare variants with moderate effects, structural variants (i.e. CNVs), genetic and gene-environment

interactions (MANOLIO *et al.* 2009). This observation further highlights the need for a robust and powerful strategy to detect pathogenic variants not identified in GWAS, and determine the biological/clinical relevance of disease-associated mutations. Importantly, a comprehensive picture of the genetic architecture in complex disorders can be obtained using a functional genomics approach (discussed in the following section), integrating genome-wide datasets on genetic variation and mutations at the DNA level, with gene expression and functional information at the RNA and protein levels.

In the work described in Chapter IV, these methods were used to characterize the complex genetic architecture of alopecia areata, and were instrumental in the identification of a novel gene cluster that plays a crucial role in disease pathogenesis.

Targeted Deep Sequencing

GWAS test for association with tag SNPs, meant to capture the variation in a particular region, but are only representative of a small subset of the variation in that region. The next step after demonstrating an association is to determine what variants are in linkage disequilibrium with the tagged SNP and could be driving association. To focus in on a particular region of genetic association or inheritance (i.e. based on linkage disequilibrium peaks), targeted deep resequencing is a suitable approach to identify both rare and common variants that exist within a selected genomic region. However, in cases where deep sequencing is not feasible, genotype-based methodologies (i.e. Sequenom genotyping, Agilent and HiSeq targeted sequencing) are used to test for known causal or associated variants.

Allelic frequencies for genetic variants can be determined through comparing their prevalence in affected to healthy cohorts as well as to the general population using

public genome database information (1000 genomes, dbSNP, NCBI, UCSC, etc.). To prioritize the biologically interesting loci for targeted deep resequencing, data obtained from other studies can provide useful information for narrowing the list of candidate regions contributing to disease pathogenesis. These sources include family-based studies (linkage, whole-exome, CNV analysis) for some complex diseases, gene expression studies (i.e. microarrays, RNA-seq), and mouse models that recapitulate the disease phenotype. More recently, these approaches have successfully identified both non-coding and coding variants in polygenic traits including various autoimmune disorders (VERLAAN *et al.* 2009; HUTCHINSON *et al.* 2014) (RIVAS *et al.* 2011; BEAUDOIN *et al.* 2013).

Whole-genome and targeted resequencing approaches provide comprehensive polymorphism and mutation discovery in human genomes. Unlike Mendelian disorders, the large majority of genomic regions in strong association with complex, polygenic traits (i.e. autoimmune diseases) lie outside of genes (intergenic) and within presumptive regulatory regions (HINDORFF *et al.* 2009; XU AND TAYLOR 2009). Functional genomics studies can be used to test the consequence of both coding and non-coding variants on gene networks by integrating data on DNA sequence, expression and genetic/protein interactions. As in rare Mendelian disorders, further functional studies (*in vitro* and *in vivo*) should be performed to test the effects of the variants and their potential contribution to disease pathogenesis.

Functional Genomic Studies

To determine the biological involvement of disease variants and their interactions at the molecular, genetic and epigenetic levels, researchers have recently begun to use

functional genomics to gain a clear and comprehensive picture of the relationship between genotype and phenotype. The Human Genome Project as well as the next-generation sequencing methods developed thereafter (discussed above) have been instrumental in the implementation of this approach to studying complex human diseases, where the ultimate goal is to validate the biological and clinical relevance of disease-associated mutations. As a follow-up to the Human Genome Project, the ENCODE (Encyclopedia of DNA Elements) Project was developed to identify all the functional regulatory elements in the human genome and currently serves as a comprehensive and indispensible resource for the scientific community (ENCODE 2004). The ENCODE and Roadmap Epigenomics Projects, as well as FANTOM5 (functional annotation of the mammalian genome 5) have facilitated the identification of novel DNA regulatory elements including gene promoters, enhancers, transcription factor binding sites, and regions of the genome subject to chromatin and histone modifications. In the context of both natural human variation and genetic disease, these data are of tremendous value in propelling the field of functional genomics, revealing how non-coding DNA can influence gene transcription and regulation, and the downstream consequences of non-coding pathogenic variants in human disease.

The functional genomics approach integrates genetic/mutational analyses with functional and regulatory information on gene transcription, translation, and protein-protein interactions in a high-throughput manner rather than a candidate-by-candidate approach. At the mRNA level, transcriptome sequencing (i.e. RNA-seq) provides information on gene expression, splicing, and discovery of transcribed SNPs or somatic mutations (MOROZOVA AND MARRA 2008). At the chromatin level, genome-wide mapping

of protein-DNA interactions can be achieved using chromatin immunoprecipitation sequencing (ChIP-seq) (RIVERA AND REN 2013). Bioinformatics analysis, coupled together with computational biology tools can be used to functionally integrate this information and provide a clear and potentially meaningful representation of the data for further interpretation and downstream analyses.

Following the identification of variants, the next step is often testing their downstream consequences on gene expression. This can be accomplished using various methods and approaches, including microarray analysis, quantitative RT-PCR (qRT-PCR) and/or RNA-sequencing. Additionally, this information can be correlated with histological/immunohistochemistry analysis of patient samples, followed by functional studies employing cell culture systems *(in vitro*) and mouse models (*in vivo*) to test a how a given coding or non-coding variant behaves and its overall contribution to disease pathogenesis.

In the work described in Chapter IV, targeted deep resequencing of a disease-associated locus, followed by functional genomics studies enabled the identification and characterization of rare, non-coding variants in alopecia areata.

Chapter I.3. The Biology of the Skin and Hair Follicle

3.A Mammalian skin and hair follicle development

The skin is the largest organ in the body, and serves as a protective barrier against injury, possessing both flexible and rigid properties governed by the types of molecules expressed in each cell that permit proper association and communication between individual cells (DESTEFANO AND CHRISTIANO 2014) (FUCHS 2007). Several different cell types of both mesenchymal and epithelial origin make up the skin: the ectodermally derived epidermis is composed of keratinocytes, and the underlying dermis consists of a heterogeneous population of mesenchymal cells that includes fibroblasts that secrete the connective tissue matrix. In the epidermis, as in all stratified squamous epithelia (i.e. esophagus and cornea), there are undifferentiated and differentiated keratinocytes expressing specific type-I (acidic) and type-II (basic) keratin intermediate filaments, as well as structural proteins required for intercellular adhesion.

Among the ectodermal appendages of the skin, the hair follicle is a miniorgan of a highly complex and dynamic structure, providing physical protection, thermoregulation, sensory perception, and cues in psychosocial and sexual communication. In mammals, hair follicle development arises through a series of reciprocal signaling events that occur between the undifferentiated epithelium and underlying mesenchyme, where gradients of activators/inhibitors regulate the patterning of hair follicles across the body. During morphogenesis, a signal emanating from the dermis instructs the epidermis to thicken and form a placode (Figure 4). At around E14.0 in the mouse, a signal from the epidermis then instructs these underlying mesodermal cells that will become the dermal papilla to condense at the site of the

newly-forming follicle (MILLAR 2002). As the epidermis proliferates, it envelops the dermal condensate, which leads to the formation and down growth of the developing hair follicle. The developing outer root sheath (ORS) (the outer most epithelial layer in the hair follicle) is continuous with the epidermis, and residing at the base of the hair follicle are matrix cells (also called transit amplifying cells), which are in direct contact with the dermal papilla (Figures 4 and 5).

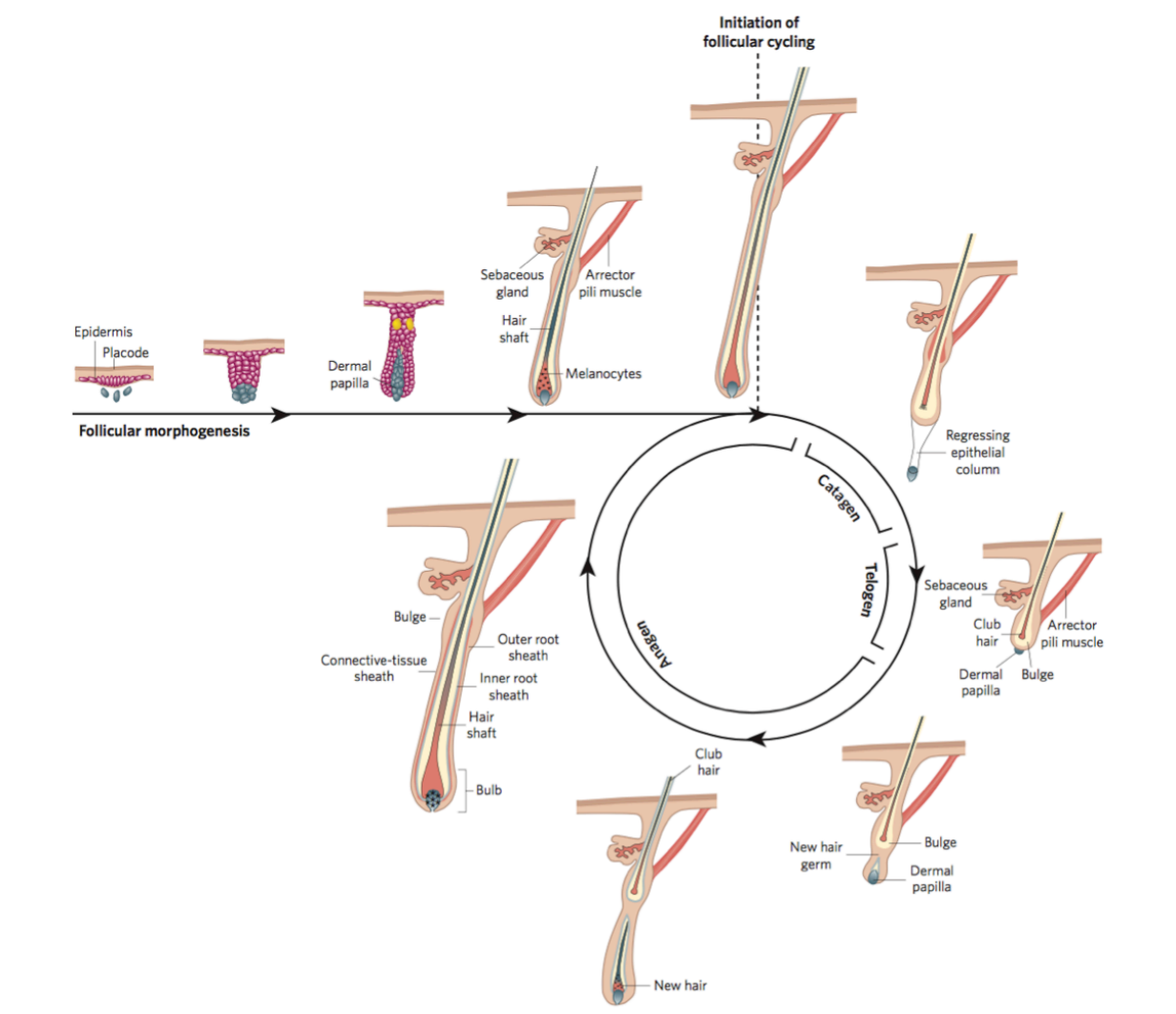


Figure 4. Hair follicle morphogenesis and cycling. Following morphogenesis, mature hairs cycle through the three main phases: anagen (growth), catagen (destructive), and telogen (resting). (Adapted from Fuchs, 2007)

During hair follicle growth, matrix cells proliferate and differentiate upwards to give rise to the three cylinders of the inner root sheath (IRS) and the hair shaft within the center of the follicle, but they have a limited proliferative capacity (Figure 5). Following exhaustion, they are replenished from a reservoir of epithelial stem cells located in the bulge, which serves a primary stem cell niche for the hair follicle and mediates the self-renewal and continuous cycling through the active growth (anagen), regression (catagen), and resting (telogen) phases (OHYAMA 2007) (FUCHS 2007) (Figure 4). Importantly, the hair follicle is an excellent system in which to study genetics due to its readily accessible nature, cycling properties, developmental pathways, diverse range of cell types (Figure 5), and widely studied spatiotemporal regulation of gene expression.

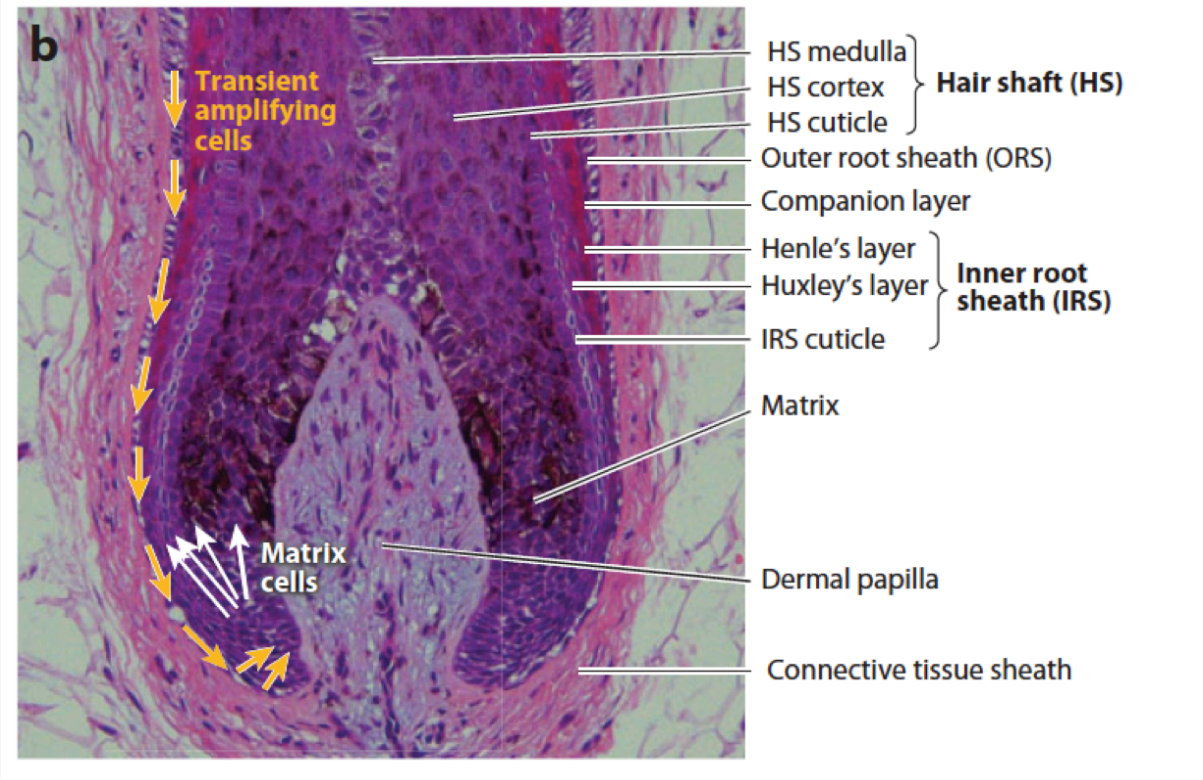


Figure 5. The human hair follicle. Hematoxylin and eosin staining of the human anagen hair follicle bulb. Transient amplifying cells derived from the bulge portion migrate to the matrix region (orange arrows), and matrix cells differentiate into distinct cell layers of the hair follicle (white arrows). (Adapted from Shimomura and Christiano, 2010)

3.B Hair patterning

Hair types in humans can be grouped into three categories based on size, cellular composition, pigmentation and androgen responsiveness (GARN 1951). Lanugo hairs are the first to form during development (appearing on the scalp by the 10th gestational week) and are uniformly long, fine and unpigmented. These hairs are shed *in utero* in a synchronized manner and are predominantly replaced by vellus hairs, which are fine and unpigmented, but shorter than lanugo hairs. Unlike the vellus hairs that cover the majority of the body, terminal hairs are the course, pigmented, and medullated hair follicles that penetrate into the fatty layer of the dermis. Terminal hair follicles first develop on the scalp, replacing the vellus intermediate-like hairs from 3-7 months of age and display characteristic features of terminal hairs by age 2, later appearing on additional sites of the body subject to androgen stimulation (i.e. sites of secondary sex characteristics) during puberty (GARN 1951).

Compared to other mammals, humans have lost most of their body hair over the course of the 90 million years of evolution. However, regional differences in hair patterning across the body found on other mammals are conserved. In both humans and lower-level mammals (i.e. *Mus musculus*), the cellular origin of the ectodermal compartment of the skin and hair follicle is non-neural and relatively uniform across the body, whereas the mesenchyme is spatially derived from distinct cellular lineages during development: the somites (dorsal dermis), lateral plate mesoderm (ventral dermis), and the neural crest and paraxial mesoderm (craniofacial dermis) (TRAN *et al.* 2010). Tissue recombination experiments performed several decades ago have revealed that during hair follicle development, the dermal mesenchyme possesses the

inductive factors responsible for initiating hair follicle growth, specifying location, size, and proper distribution pattern of the developing hair follicles (SENNETT AND RENDL 2012). Interestingly, studies in humans and mice have demonstrated that dermal fibroblasts derived from different anatomic positions possess strikingly distinct gene expression profiles that are correlated with the anterior-posterior, proximal-distal, and dermal-nondermal axes (RINN *et al.* 2006; RINN *et al.* 2008).

As the human hair follicle cycles through anagen, catagen, and telogen, it is not synchronized with its neighboring follicles or with those at different body sites (in contrast to the synchronization in mouse hair follicles). Thus, depending on the location and type of hair produced, the hair cycle can last from months up to several years. On the scalp, hairs remain in anagen for 2-6 years, in catagen for 2-3 weeks, and telogen for ~3 months, whereas the vellus hairs covering the body are primarily in telogen, only spending 6-12 weeks in anagen. It is likely that differences in mesenchymal cell origin and initial cell fate specification contribute to the visible patterning and regionalization differences of hair follicles found on the scalp, face, and body surface.

Following hair patterning, hair follicles can retain the identity and characteristics of the donor site when transplanted to a recipient site, a phenomenon called donor dominance, is observed in hair transplantation on the scalp (back of the head to the front), as well as scalp or facial hairs (eyebrows) transplanted to other sites of the body (i.e. the leg) (ORENTREICH 1959). Importantly, this remarkable property forms the basis for hair transplantation in individuals with androgenetic alopecia (male-pattern baldness), where hairs relocated from the back of the scalp to the front survive for decades (ORENTREICH 1959; BARRETO 1998; LEE *et al.* 1999; GANDELMAN AND EPSTEIN

2004). While it is known that certain regions of the scalp (such as the occipital fringe) remain refractive to the effects of androgens while others are susceptible (front and top), the molecular mechanisms governing this phenomenon have not been studied.

3.C Mouse Models for Hair Follicle Morphogenesis and Cycling

While there are several notable differences between human and mouse hair follicles (i.e. hair types, cycling properties, etc), the mouse as a model has proven useful in identifying the genes and pathways that control hair follicle morphogenesis and cycling (NAKAMURA *et al.* 2013). Since the molecular and biological pathways underlying hair follicle growth have been well characterized in mice, most of what we know about the controls of hair growth, size, patterning has been determined by the use of spontaneous and genetically-engineered mouse models exhibiting hair follicle defects (NAKAMURA *et al.* 2013). For example, *angora* mice null for the *Fgf5* gene have abnormally long hair, due to a prolonged anagen phase (HEBERT *et al.* 1994). Interestingly, mutations in the *FGF5* gene have also been identified in *angora* rabbits, sheep, cats, goats and most recently, the Christiano lab reported *FGF5* mutations in humans with trichomegaly (long eyelashes) (HIGGINS *et al.* 2014), highlighting the critical and conserved role for this growth factor in regulating hair length.

Over the past few decades, studies in mouse models have revealed that a wide variety of well-conserved growth factors and molecules involved in cell signaling, adhesion and communication are crucial for the development and dynamic maintenance of the skin and hair follicle. The genes regulating the development of hair follicles are very similar (and in some cases, identical) to those regulating the formation of other ectodermal appendages (i.e. teeth, mammary gland), such that mutations in these

genes produce pleiotropic effects (SCHMIDT-ULLRICH AND PAUS 2005; NAKAMURA *et al.* 2013). Many studies have demonstrated a key role for the Wnt/β-catenin signaling pathway during hair follicle patterning and morphogenesis, where forced expression of β-catenin in the mouse epithelium leads to an overall increase in the number in hair follicles, an acceleration in morphogenesis and growth, and the production of hair-derived tumors (GAT *et al.* 1998). Conversely, ectopic expression of the Wnt inhibitor, Dickkopf (Dkk1) in epidermal keratinocytes abrogates placode formation and patterning (ANDL *et al.* 2002).

In addition to the Wnt/β-catenin signaling pathway, components of the Bmp, Shh, Eda/EdaR, Fgf, and Pdgf signaling pathways, as well as several transcription factors (Sox9, Hoxc13, Msx1/2, Lef1, Gata3, etc.) possess essentials roles in morphogenesis and hair growth (SCHMIDT-ULLRICH AND PAUS 2005; NAKAMURA *et al.* 2013). These pathways are known to converge, since mouse models with targeted or spontaneous mutations in these genes, depending on the developmental time point, can act both upstream and downstream of one another (i.e. Wnt signaling controls and is controlled by the Eda/EdaR/NF-KB signaling pathway during hair follicle induction). Importantly, positive and negative feedback regulatory loops are required for augmenting or attenuating gene expression levels, and this spatiotemporal regulation is required for proper development of the skin and hair follicle.

3.D Genetic Mutations in Human Hereditary Hair Disorders

In humans, mutations in key signaling molecules, transcription factors, structural proteins, and transporters have been identified in several Mendelian ectodermal dysplasias, or abnormal development of the hair, skin, nails, teeth and/or eccrine glands

(SHIMOMURA AND CHRISTIANO 2010). Notably, the phenotypes in human genetic skin and hair disorders are often recapitulated in the mouse models harboring mutations in these same genes (SCHMIDT-ULLRICH AND PAUS 2005; NAKAMURA *et al.* 2013). Humans with autosomal recessive odonto-onycho-dermal dysplasia, characterized by hypotrichosis, hypodontia, nail dystrophy and keratoderma, possess loss-of-function mutations in the *WNT10A* gene (ADAIMY *et al.* 2007). Similarly, in hereditary hypotrichosis simplex, an autosomal dominant nonsyndromic disorder of hair loss, the Christiano lab previously reported heterozygous mutations in the *APCDD1* gene, encoding a membrane-bound repressor of the WNT/*β*-catenin signaling pathway, in large families from Pakistan (SHIMOMURA *et al.* 2010a). In addition to WNT signaling, proteins of the TNF pathway, EDA, EDAR, and EDARADD regulate hair follicle patterning and morphogenesis, where mutations in these proteins cause both autosomal and X-linked forms of hypohidrotic ectodermal dysplasia (HED), characterized by abnormal development of the hair, skin, nails, teeth and eccrine glands (CLUZEAU *et al.* 2011). Moreover, the human HED phenotype recapitulates that observed in four spontaneous mouse models that harbor mutations in the homologous genes (*Tabby/Sleek/downless/crinkled,* named after the phenotypes) (SHIMOMURA AND CHRISTIANO 2010).

Signaling pathways that regulate lipid metabolism (i.e. biosynthetic enzymes) and ion transport also play an important role in regulating hair growth and the overall structural integrity of the hair follicle and shaft. Mutations have been identified in the *LPAR6* (also known as *P2RY5*, encoding an orphan G-protein coupled receptor for LPA), and *LIPH* (lipase H) genes in humans with LAH and autosomal recessive woolly hair syndromes (KAZANTSEVA *et al.* 2006; SHIMOMURA *et al.* 2008; SHIMOMURA *et al.*

2009), where the *Liph* knockout mouse phenotype recapitulates the human condition (INOUE *et al.* 2011). Interestingly, these genes are all expressed in the IRS of human hair follicles and perfectly overlap with the expression of the *KRT71* and *KRT74* genes that underlie autosomal dominant forms of woolly hair/hypotrichosis (SHIMOMURA *et al.* 2008; FUJIMOTO *et al.* 2012). Furthermore, it is known that *LIPH* encodes a membrane-associated phosphatidic acid-selective phospholipase that produces 2–acyl lysophosphatidic acid (LPA) from phosphatidic acid, and that phosphatidic acid promotes hair growth, highlighting the critical role for these lipid mediators in the overall integrity of the hair follicle (SONODA *et al.* 2002; PASTERNACK *et al.* 2008).

Several genes encoding transcription factors required for patterning and differentiation of hair follicle cells are also mutated in hereditary hair disorders. Classic examples include *p63* in several autosomal dominant diseases (i.e. ectodermal dysplasia) (CELLI *et al.* 1999), the zinc-finger transcription factor/histone demethylase *Hairless (HR)* that regulates catagen and is mutated in atrichia with papular lesions (AHMAD *et al.* 1998; PANTELEYEV *et al.* 1999), and *FOXN1*, a forkhead box transcription factor expressed in the matrix and hair shaft whose loss-of-function causes T cell immunodeficiency, congenital alopecia (hair loss), and nail dystrophy (FRANK *et al.* 1999; SCHLAKE *et al.* 2000).

In addition to transcription factors, a great number of mutations have been identified in genes encoding structural proteins and cellular adhesion molecules (i.e. components of the desmosome) in hereditary hair disorders, where mutations in the *KRT81, KRT83,* and *KRT86* genes underlie monilethrix (fragility of scalp hairs), and *KRT74* and *KRT71* mutations cause autosomal dominant woolly hair syndrome and

hypotrichosis (WINTER *et al.* 1997; CHIEN *et al.* 2006; SHIMOMURA *et al.* 2010b). Similarly, we previously identified mutations in the *DSG4* gene in patients with localized autosomal recessive hypotrichosis (LAH), and in conjunction with comparative genomics studies in the *lanceolate hair* mouse, defined the critical role for *DSG4*, a suprabasally expressed desmoglein, in the skin and hair follicle (KLJUIC *et al.* 2003). Importantly, this was one of the first reports of positional cloning that was accelerated by the Human Genome Project (2003) because *DSG4* had not been identified by previous cloning or sequencing methods.

Whereas genetic mutations in structural and signaling proteins result in hair anomalies, other factors including hormones, nutrients, and environmental stress can perturb hair growth and in some cases, phenocopy a hereditary disorder. Human patients with hypo- and hyperthyroidism experience hair loss, and mice deficient for thyroid hormone receptors TRα1 and TRβ have impaired epidermal proliferation, hair growth and wound healing as a result of defective mobilization of stem cells out of the niche (CONTRERAS-JURADO *et al.* 2015). Additionally, neurotransmitters, growth factors, cytokines and steroids are all molecules delivered to the hair follicle via vasculature and/or neighboring cell types and are reported to influence the dynamics of hair growth.

Hair follicle morphogenesis and growth are tightly regulated processes, where even minor perturbations lead to abnormalities in size, pattern, structure and the number of hair follicles. Importantly, the many genetic studies performed over the past several decades have taught us a great deal about the complexity of the skin and hair follicle, and the various classes of molecules that are responsible for maintaining their overall structural integrity.

Chapter I.4. Inherited Hypertrichosis

Inherited hypertrichoses are very rare human disorders characterized by excessive hair growth that does not depend on androgen stimulation and is independent of age, gender, and ethnicity (BEIGHTON 1970). Hypertrichosis syndromes fall under the larger umbrella of ectodermal dysplasias, or abnormal development of the hair, skin, nails, teeth and/or eccrine glands, and are often associated with additional anomalies including gingival hyperplasia, deafness, cardiomegaly, and bone abnormalities (GARCIA-CRUZ *et al.* 2002). Dating back to the Middle Ages, individuals with hypertrichosis became exhibitionists for the purposes of entertainment and art, as the striking image of an individual completely covered in hair has long fascinated humans, particularly evolutionary geneticists (Figure 6). These conditions are extremely rare, with only approximately 60 reported cases to date, and it has been suggested that inherited hypertrichoses represent examples of atavisms, or the recurrence of an ancestral phenotype, where the genes that promote a full coat of hair in other mammals and were silenced throughout evolution have become reactivated in human hypertrichosis, invoking unusual genetic mechanisms to explain their occurrence (HALL 1984; CANTU AND RUIZ 1985).

The excessive hair overgrowth phenotype of hypertrichosis can be produced from drug reactions, hormone imbalances and toxic environmental effects. Similarly, there are several syndromes with congenital localized hypertrichosis as an associated feature, including Cantú syndrome (OMIM 239850) (CANTU *et al.* 1982a; CANTU *et al.* 1982b), a condition first described in 1982 that is characterized by hypertrichosis, coarse facial features, cardiac hypertrophy and osteochondrodysplasia, and Coffin-Siris

syndrome (OMIM 135900) (QAZI *et al.* 1990), a rare autosomal dominant condition of mental retardation, craniofacial anomalies, hypertrichosis and other ectodermal dysplasias. Importantly, several forms of hypertrichosis as an inherited, isolated disorder have been described, where the excessive hair overgrowth phenotype is the primary clinical feature and is most often generalized rather than localized. These include hypertrichosis universalis congenita, Ambras type (OMIM 145701), hypertrichosis lanuginosa congenita (OMIM 145700), X-linked hypertrichosis (OMIM 307150), and congenital generalized hypertrichosis terminalis with or without gingival hyperplasia (OMIM 135400) (BEIGHTON 1970) (Figure 6).

Figure 6. Inherited Hypertrichosis. Clinical photos of Ambras syndrome, CGHT, and X-linked hypertrichosis patients. (Adapted from Baumeister et al. 1993, Sun et al. 2009, DeStefano et al. 2013, Zhu et al. 2011, and DeStefano et al. 2014)

4.A Structural and Copy Number Variants Associated with Hypertrichosis

Up until the past decade, the underlying genetic basis of these syndromes remained elusive, but recent studies from the Christiano lab and others have identified structural defects, copy number variants, and position effects associated with the autosomal dominant, sporadic and X-linked forms (discussed in Chapter 1.5). In hypertrichosis congenita, Ambras type (a form of hypertrichosis that affects the shoulders, forehead, nose, cheeks, eyelids, and ears), the Christiano lab defined a candidate interval on chromosome 8q containing the *TRPS1* gene based on the cytogenetic breakpoints in three patients (TADIN *et al.* 2001; FANTAUZZO *et al.* 2008). *TRPS1* was deleted in a patient with an 8q23 chromosomal rearrangement, and its expression was significantly downregulated in another patient with an inversion breakpoint 7.3Mb downstream of *TRPS1* (FANTAUZZO *et al.* 2008).

In the autosomal dominant form of congenital generalized hypertrichosis terminalis (CGHT), the Christiano lab identified a series of microduplications within a 2.4 Mb region on chromosome 17q24.2-24.3 ~1 Mb upstream of the SOX9 gene (FANTAUZZO *et al.* 2012). In an independent report, CNVs were identified in three CGHT families and one sporadic case, where the underlying genetic lesions were three distinct microdeletions in the familial cases, as well as one inverted duplication in the sporadic case (SUN *et al.* 2009). Taken together, these results strongly implicate the chr17q24.2-24.3 region in CGHT pathogenesis, providing an allelic series of five reported CNVs.

In Chapter III, I further studied the chr17q24.2-24.3 region to identify genetic mechanisms underlying human hair growth in two independent CGHT cases (one sporadic, one familial).

4.B Position Effects on Genes Regulating Hair Follicle Development

Position effects on single genes as a result of large chromosomal rearrangements or copy number variants have been reported in several ectodermal dysplasias, including hypertrichosis syndromes. Despite the pleiotrophic phenotypes of these syndromes and their reported association with large chromosomal rearrangements and copy number variants, the causative genetic mechanism underlying human hypertrichoses remained elusive.

Within the past five years, position effects as the underlying genetic basis of these syndromes have been reported on the *TRPS1* and *SOX9* genes that reside ~1 Mb to as distant as 7.3 Mb from these chromosomal rearrangements in hypertrichosis congenita, Ambras type and congenital generalized hypertrichosis terminalis (CGHT), respectively (FANTAUZZO *et al.* 2008; FANTAUZZO *et al.* 2012). In humans with autosomal dominant and sporadic forms of hypertrichosis, the Christiano lab identified position effects on *SOX9* as a result of a 1.2 Mb deletion upstream of the gene, and observed reduced *SOX9* expression at the mRNA and protein levels throughout patient hair follicles (FANTAUZZO *et al.* 2012). Several reports have demonstrated a crucial role for SOX9 in specifying hair follicle stem cells, where the conditional ablation of *Sox9* in mice results in depletion of the stem cell niche and progressive hair loss (VIDAL *et al.* 2005; NOWAK *et al.* 2008; KADAJA *et al.* 2014). This finding, though seemingly contradictory to the human phenotype of excessive hair growth from decreased *SOX9* expression, highlights the paradoxical nature of position effects in generating distinct phenotypes from that produced by intragenic loss-of-function mutations.

In the X-linked forms of hypertrichosis, the Christiano lab and other groups have

reported linkage to chromosome Xq24-27 in two independent Mexican families (FIGUERA *et al.* 1995; TADIN-STRAPPS *et al.* 2003). More recently, large interchromosomal insertions were been reported in one of these Mexican families and a Chinese family containing linkage to chromosome Xq24-27, where the insertions in both families occurred at the exact same Xq27.1 extragenic palindrome site (ZHU *et al.* 2011). Importantly, the sequences contained within each insertion differed between the families, suggesting that the presence rather than the content was responsible for mediating the excessive hair overgrowth phenotype.

In Chapter II, I investigated the underlying genetic mechanism in the Mexican family with X-linked congenital generalized hypertrichosis in which the Christiano lab previously identified linkage to chr24-27 over a decade ago.

4.C Intragenic mutations

The genetic basis of hypertrichosis syndromes is largely composed of structural anomalies and CNVs. Prior to the work described in Chapter III, there were no reported intragenic mutations underlying CGHT cases. However, intragenic mutations have been reported in rare genetic conditions manifesting hypertrichosis as an associated (but not primary) clinical feature, including Coffin-Siris syndrome, Cantú syndrome, Schinzel-Giedion syndrome, and Cornelia de Lange syndrome. In patients with Coffin-Siris syndrome, a rare autosomal dominant condition of mental retardation, craniofacial anomalies and ectodermal dysplasias (including hypertrichosis), whole-exome sequencing identified fifteen mutations in genes encoding subunits of the SWI/SNF chromatin remodeling complex (TSURUSAKI *et al.* 2012). Over the past few years, several independent groups have studied the genetic basis of Cantú syndrome, and

whole-exome sequencing in families and unrelated cases identified intragenic, dominant gain-of-function mutations in the *ABCC9* gene (HARAKALOVA *et al.* 2012; VAN BON *et al.* 2012). *ABCC9* encodes the SUR2 protein component of the ATP-sensitive potassium channel complex. Minoxidil (Rogaine) is an agonist of this potassium channel and directly binds the SUR2 subunit (SCHWANSTECHER *et al.* 1998). Interestingly, the phenotypic similarities between Cantú syndrome patients and individuals with side effects from Minoxidil suggest that the dominant mutations in *ABCC9* cause constitutive channel opening (VAN BON *et al.* 2012). Importantly, this observation further supports a role for *ABCC9* in regulating human hair growth, highlighting the significance of ion transport in the hair follicle.

In patients with Schinzel-Giedion midface retraction syndrome, characterized by mental retardation, multiple congenital skeletal anomies, hypertrichosis and distinctive facial features, heterozygous de novo mutations in the *SETBP1* gene were identified by whole-exome sequencing (HOISCHEN *et al.* 2010). Similarly, mutations in the *NIPBL* gene have been identified in the autosomal cases of Cornelia de Lange syndrome type I, where patients possess facial dysmorphism, hypertrichosis, mental retardation and several craniofacial anomalies (PIE *et al.* 2010). In the X-linked forms, mutations have been identified in other components of the cohesin complex (*SMC1A, SMC3*, and *RAD21*) and in the *HDAC8* gene (ROHATGI *et al.* 2010) (MUSIO *et al.* 2006).

In Chapter III, I studied the genetic basis of autosomal recessive CGHT using whole-exome sequencing and characterized a novel transporter in human hair growth.

4.D Mouse models

Although little is known about the genes and mechanisms perturbed in hereditary

hypertrichosis syndromes, mouse models have proved useful for recapitulating the phenotype and identifying genetic pathways that regulate hair growth. As mentioned in Chapter I.3, the *angora* mouse harboring a mutation in *Fgf5* has a prolonged anagen phase, leading to a striking increase in length of the pelage hairs (HEBERT *et al.* 1994). Similarly, forced expression of *Noggin*, a BMP antagonist, leads to the production of enlarged anagen hair follicles (SHAROV *et al.* 2006), and targeted overexpression of *β*-catenin in the epidermis during morphogenesis leads to an overall increase in the number of hairs induced, and accelerated morphogenesis and growth (GAT *et al.* 1998).

The *Koala* mutant mouse is a model for hypertrichosis, containing a chromosomal inversion whose proximal breakpoint maps 791 bp upstream of *Trps1* (Fantauzzo et al. 2008). Remarkably, the position effect the Christiano lab identified on *Trps1* associated with Ambras syndrome was recapitulated in the *Koala* mouse, which demonstrated a marked decrease in *Trps1* expression at the sites of pathology for the phenotype (Fantauzzo et al. 2008). Further functional studies performed in the lab demonstrated a role for TRPS1 as a novel regulator of the Wnt signaling pathway. *SOX9* was identified as a direct transcriptional target gene of TRPS1, where these genes form an important feedback loop with SHH to regulate epithelial proliferation in the hair follicle (FANTAUZZO AND CHRISTIANO 2012; FANTAUZZO *et al.* 2012).

Importantly, functional studies performed in mouse models, in conjunction with a thorough analysis of human patient samples suggests three mechanisms by which the excessive hair overgrowth phenotype may arise: an increased duration of anagen, an overall increase in the number of hair follicles during induction, and a vellus-to-terminal transformation, where fine, unpigmented hairs are stimulated to grow larger and extend

deep into the dermis, become pigmented and form a medulla. In the context of hypertrichosis, it is likely that human patients possess a combination of these three defects to produce the excessive hair overgrowth phenotype.

In Chapter II, I investigated genomic effects involving structural variants and position effects on human hair growth, and in both Chapters II and III, discuss these genetic mechanisms underlying both X-linked and autosomal forms of hypertrichosis.

5. The Genetic Basis of Alopecia Areata

5.A Epidemiology and Immunopathology

The human hair follicle is considered an immune privileged site, tolerating antigen production without inducing an inflammatory response (ITO *et al.* 2008b). Hair follicle immune privilege is accomplished through the downregulation of MHC class I molecules that are involved in antigen presentation, active repression of ligands that stimulate natural killer (NK) cell activation, and the secretion of immunosuppressive molecules (i.e. TGFβ1 and β2, and macrophage migration inhibitory factor, MIF) that allow the hair follicle to evade immunosurveillance (GILHAR *et al.* 2012).

Alopecia areata (AA) is a complex autoimmune disease characterized by disfiguring hair loss due to the collapse of immune privilege in the hair follicle. Among all autoimmune diseases, AA is one of the most common (lifetime risk of 1.7%), affecting approximately 5.3 million people of both genders and of varying age and race (PETUKHOVA AND CHRISTIANO 2013). Moreover, AA is one of the most burdensome chronic skin diseases with an unpredictable onset and relapse period, producing psychologically devastating effects on an individual's self-image and self-esteem. The

onset of AA has been described as the sudden whitening of hair, as the target of autoimmune attack is the growing (anagen), pigmented follicle, resulting in the loss of patches of hair (patchy AA), all scalp hair (alopecia totalis), or complete body hair (alopecia universalis) (GILHAR AND KALISH 2006) (Figure 7, A-D).

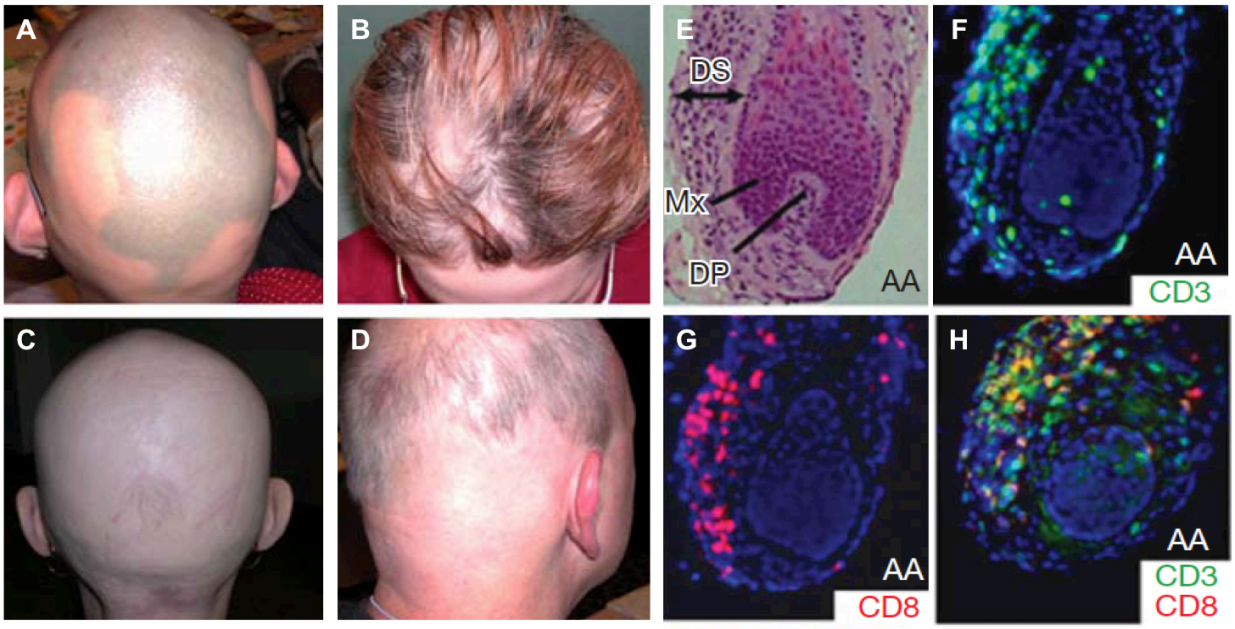


Figure 7. Clinical Immunopathology of Alopecia Areata. A-D. AA patients with patchy AA **(A)**, in the remission/regrowth phase **(B, D),** and with alopecia universalis. **E-H.** AA hair follicles (Hematoxylin and eosin stain, **A**) infiltrated with CD3+ **(F, H)** and CD8+ **(G, H)** T cells. (Adapted from Petukhova et al. 2010)

During active disease, the hair cycle is disrupted by the early onset into catagen, leading to a shift in the ratio of anagen to telogen and catagen hair follicles. The immunopathology of human AA is characterized by the infiltration of CD3+ and CD8+ T lymphocytes at the base of the hair follicle (resembling a 'swarm of bees') (Figure 7, E-H), where the immune infiltrate is composed of cytotoxic T cells, natural killer (NK) cells as well as activated T-helper cells (GILHAR AND KALISH 2006). The triggering event in active disease is thought to be due in part to the loss of immune privilege, and the pathogenicity of this immune infiltrate has been demonstrated in human xenograft models, where lesional human lymphocytes transferred to immunocompromised SCID

mice bearing a human scalp skin graft were able to induce disease (GILHAR *et al.* 1998). Although T lymphocyte-mediated autoimmune attack causes destruction of the AA hair follicle, there is no permanent organ damage as the stem cell compartment is spared (similar to the destruction in catagen), thus making hair regrowth possible.

5.B Genetic studies in family-based cohorts

Evidence supporting a genetic basis for AA includes the observed heritability in first-degree relatives (10-fold increased risk) (VAN DER STEEN *et al.* 1992; MCDONAGH AND TAZI-AHNINI 2002), concordance rates in twin studies (55% concordance in monozygotic (MZ) twins; 0% in dizygotic (DZ) twins) (JACKOW *et al.* 1998) and family-based linkage studies (MARTINEZ-MIR *et al.* 2003). While the majority of the genetic studies in AA were conducted using candidate association analysis, the Christiano lab conducted the first genome-wide scan for linkage with AA, and identified evidence for linkage on chromosomes 6, 10, 16 and 18 that have been replicated in subsequent studies (MARTINEZ-MIR *et al.* 2007). The *HLA* region on chromosome 6p possessed the strongest signal in both association studies and linkage analysis, pointing to specific *HLA-DQB* and *HLA-DR* alleles that conferred risk. Outside the *HLA* region, a significant peak on chromosome 6q25.1 was present in a region not previously associated with an autoimmune disease, and this association is discussed in following section.

In a subsequent study, the *MICA* locus was subsequently identified in an extended haplotype of the *HLA* region in strong association with AA, where the *MICA* (MHC class I chain-related gene A) gene encodes a ligand for the cytotoxic T cell receptor, NKG2D (BARAHMANI *et al.* 2006). NKG2D is activated in the presence of a stressed cell emitting a “danger signal” upregulating ligand expression, and this

ultimately leads to cellular destruction (CAILLAT-ZUCMAN 2006). The NKG2D signaling pathway in cytotoxicity has been well described in the context of viral infection, autoimmunity as well as cancer (GONZALEZ *et al.* 2006). The association of MICA/NKG2D signaling with autoimmune disease was previously known, implicated in the pathology of type I diabetes and rheumatoid arthritis. In the context of AA, MICA expression is upregulated in AA lesional hair follicles, a finding that together with the strong genetic association provides evidence for the NKG2D signaling pathway in AA disease pathogenesis (BARAHMANI *et al.* 2006; ITO *et al.* 2008a).

5.C GWAS

Several family-based studies have been conducted to identify susceptibility loci associated or in LD with AA, however, small sample sizes and candidate-based approaches limited their informative value. Thus, a more robust and unbiased approach using population-based studies was required to elucidate the key pathogenic risk variants underlying this complex disease. The Christiano lab conducted the first GWAS for AA using 1,054 cases and 3,278 controls, and identified eight regions of the genome significantly associated with AA that clustered into discrete LD blocks. Among the associated regions were the *HLA, ULBP3/6, CTLA4, IL-2/IL-21, IL-2RA, IKZF4, PRDX5*, and *STX17* loci, many of which have been previously implicated in other autoimmune diseases (including T1D, RA, celiac disease) (PETUKHOVA 2010) and were recently replicated in our meta-GWAS analysis (BETZ *et al.* 2015). In Chapter IV.2, I performed functional studies on novel genes associated with AA that were identified in our recent meta-GWAS analysis (BETZ *et al.* 2015).

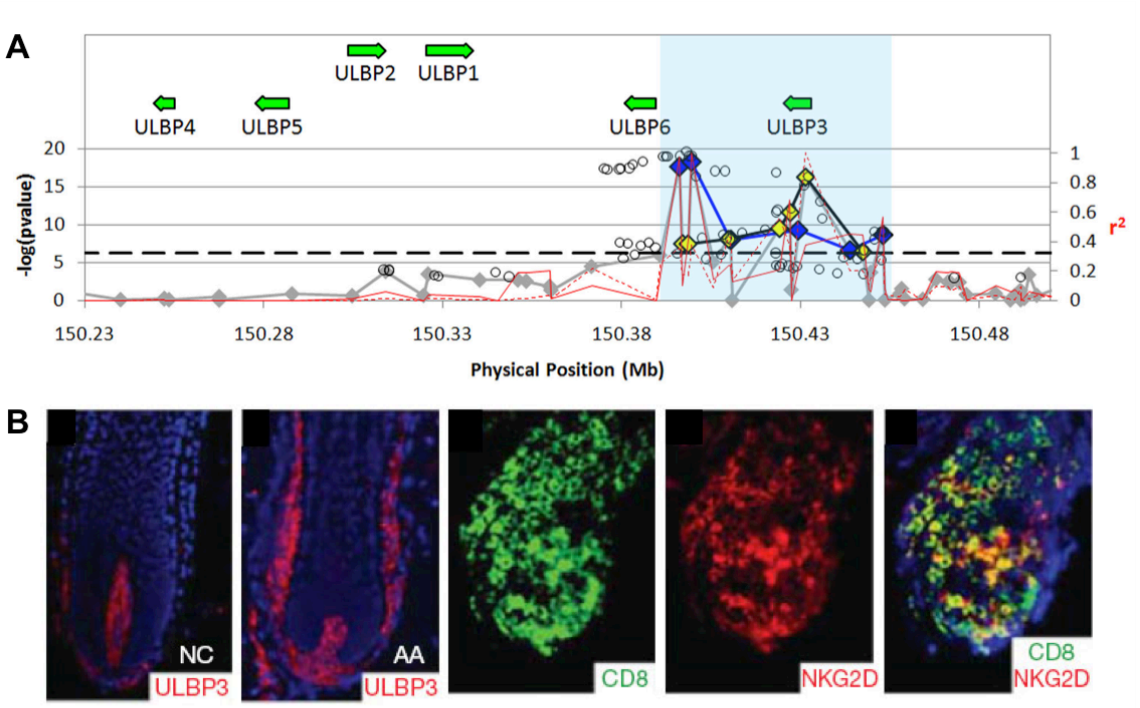


Figure 8. GWAS implicates *ULBP3/6* in AA disease pathogenesis. A. Genomic view of *ULBP3/6* association with AA (P = 4.49 x 10^{-19}); yellow and blue lines indicate different haplotypes with tag SNPs. ***B.*** ULBP3 upregulation in AA lesional hair follicles surrounded by CD8+ T cells expressing the NKG2D receptor. (Adapted from Petukhova et al. 2010)

Remarkably, the *ULBP3/6* locus on chromosome 6q25.1 possessed the highest association signal (P = 4.49 x 10^{-19}) with AA outside the *HLA*, and the significance of this association increased in our replication study (P = 5.9 x 10^{-24}) (BETZ *et al.* 2015). The *ULBP* (cytomegalovirus UL16-binding protein) genes 1-6 reside in a 180-kilobase MHC class I-related cluster (Figure 8A) and encode ligands for the activating cytotoxic T cell receptor, NKG2D (CAO AND HE 2004). While allelic variants and polymorphisms in other NKG2D ligands (i.e. MICA/MICB) have been previously associated with AA (BARAHMANI *et al.* 2006; ITO *et al.* 2008a), the GWAS study conducted in the Christiano lab is the first to implicate the *ULBP* gene cluster in any human disease. The statistical association of *ULBP3/6* with AA was biologically validated by immunofluorescence

staining, whereby ULBP3 expression was strongly upregulated in lesional hair follicles, and the T cell infiltrate at the base of the hair follicle was found to contain NKG2D+ CD8+ T cells (PETUKHOVA *et al.* 2010) (Figure 8B). Importantly, subsequent studies in mouse models have demonstrated that inducible overexpression of *Rae1 (ULBP* homologue) in the mouse epidermis, pulmonary epithelium, and β-islet cells causes infiltration of cytotoxic T cells similar to what is observed in autoimmune disease (MARKIEWICZ *et al.* 2012) (STRID *et al.* 2008; BORCHERS *et al.* 2009).

These immunological findings and the strong genetic association of *ULBP3/6* with human AA provide evidence for a role for the *ULBP*-NKG2D signaling axis, yet the precise mechanism(s) by which pathogenic NKG2D ligand variants contribute to autoimmunity in the hair follicle remained unknown. Interestingly, the majority of susceptibility variants associated with complex genetic disorders identified by GWAS are intergenic, residing within non-coding DNA and presumptive regulatory regions (MANOLIO *et al.* 2009; WANG *et al.* 2010). Moreover, the emerging picture of the genetic architecture underlying complex disease supports a role for both common and rare variants that drive associations and contribute to the "missing heritability" (MANOLIO *et al.* 2009). Considering the loss of immune privilege in the AA hair follicle and ectopic expression of key immune-related genes, I postulated that non-coding variants at the *ULBP3/6* locus contribute to the marked upregulation of these ligands in AA.

5.D C3H/HeJ mouse model for alopecia areata

An additional line of evidence in support of a genetic basis for AA comes from the study of animal models. The C3H/HeJ inbred strain spontaneously develops AA (~20% rate),

however, the disease can be induced using a well-established efficient skin grafting model (95-100% reliable), allowing us to study interventions in the context of both disease prevention and reversal (SUNDBERG *et al.* 2004) (MCELWEE *et al.* 1998). Genetic linkage studies in the C3H/HeJ strain have identified four susceptibility loci on chromosomes 8 (*Alaa3*), 9 (*Alaa2*), 15 (*Alaa4*), and 17 (*Alaa1*), where *Alaa1* corresponds to the human *HLA* region (SUNDBERG *et al.* 2004). Although the regions of genetic association differ between the human and mouse disease (excluding the *HLA* locus), the C3H/HeJ mouse model recapitulates many pathological features of adult AA, including the infiltration of T lymphocytes at the base of the growing hair follicle.

Whereas in other autoimmune diseases the end organ tends to be irreparably destroyed, the skin and hair follicles remain accessible and amenable for histological and molecular studies in AA. Thus, in human AA patients and the mouse model, we can isolate the affected target tissue to interrogate the molecular pathology both prior to and during disease pathogenesis. This feature provides us with a tremendous advantage, allowing us to dissect the transcriptional network and crosstalk between the target hair follicle and effector T cells. The trajectory and approaches we used to accomplish this ultimately led to the development of therapeutic interventions in the mouse model and now in human AA patients (Chapter IV.3).

In Chapter IV.3, I aimed to identify and characterize genes whose mRNA levels were dysregulated in the mouse model using microarray expression profiling and immunohistological analysis. In Chapter IV.4, I focused my studies on the *ULBP3/6* locus given its strong genetic association with AA and role of NKG2D signaling in disease pathogenesis. In this work, I investigated the mechanisms regulating *ULBP3/6*

expression and characterized novel susceptibility variants using functional genomics.

6. Long-Range Interactions Controlling Gene Expression

In addition to position effects (discussed in Chapter I.5), changes in the local chromatin structure and long-range interactions between enhancer elements and gene promoters play a critical role in orchestrating the spatiotemporal expression and repression of genes. These changes are accompanied by a reorganization of the nuclear architecture and the formation of active chromatin hubs engaged in both intra- and interchromosomal looping interactions. Over the past few years, several methods have been developed to detect and capture long-range interactions, including the Chromosome Conformation Capture technologies (3C, 4C, 5C, Hi-C) (DOSTIE *et al.* 2006; SIMONIS *et al.* 2006; ZHAO *et al.* 2006), ChIA-PET (Chromatin Interactions Analysis with Paired-End Tag Sequencing) (LI *et al.* 2010) for interactions bound by a protein of interest, ChIP-loop (HORIKE *et al.* 2005), a method that combines ChIP and 3C, as well as the protein-tethered *Dam* methylase technique, where methylated regions can be attributed to chromatin looping (VAN STEENSEL AND HENIKOFF 2000). In conjunction with these approaches, FISH at a high resolution (3D FISH) can be used to detect and validate the juxtaposition of distinct genomic regions (OSBORNE *et al.* 2004).

Long-range interactions have been observed in several cell types and biological processes requiring spatiotemporal control of gene expression. Depending on the context, *cis* acting regulatory elements (i.e. enhancers and repressors/insulators) can be located and act in close proximity to the gene (i.e. intronic region) or exert effects over very large distances (several Mb). Allele-specific long-range interactions have been described at the *H19/IGF2* imprinting control region (ICR) during imprinting, where

parental-specific expression of the *H19* and *IGF2* alleles is required for normal development. The CCCTC-binding factor, CTCF, has been reported to act as an insulator in preventing interactions between the *H19* enhancers and *IGF2* promoter on the maternal allele, (LI *et al.* 2008; PHILLIPS AND CORCES 2009; CUNNINGHAM *et al.* 2010; GUIBERT *et al.* 2012), where its occupancy at the ICR is inversely correlated with methylation status. Importantly, disrupted transcriptional regulation of imprinted genes can lead to severe defects, and has often been reported to cause human imprinting disorders (i.e. Beckwith-Wiedemann syndrome and chr11p15.5) (NATIVIO *et al.* 2011).

Long-range interactions in the immune system

Beyond imprinting, long-range interactions and a role for CTCF have been described in the immune system during lineage commitment, the generation of antigen diversity, as well during an immune response (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS *et al.* 2013; SHARAF *et al.* 2014). Several immune-related genes are exclusively expressed from one allele (monoallelic expression) to ensure uniformity in the recognition of molecules in the context of an immune response (i.e. antigen recognition). Over the past decade, studies have begun to uncover the complex transcriptional regulation of the highly polymorphic MHC Class I and II genes at the *HLA* locus, where the binding of CTCF to conserved insulator elements in between the *HLA-DR* and *HLA-DQ* gene clusters physically links the MHC promoters to distal enhancer elements, promoting an active configuration of transcription (MAJUMDER *et al.* 2008; MAJUMDER AND BOSS 2010; GILLEN AND HARRIS 2011). Similarly, CTCF-mediated chromatin remodeling and long-range interactions have been reported at the *IFNG* and Th2 *IL4* loci during T cell differentiation/maturation, as well as at the *IL-1/IL-*

36/IL-37 gene cluster during the activation of monocytes (LEE *et al.* 2005; SHARAF *et al.* 2014). Recent studies have demonstrated the presence of lineage-specific super enhancers (SEs) at well-characterized immune loci that engage in long-range interactions, where these SE boundaries are demarcated by CTCF binding (HNISZ *et al.* 2013; FARH *et al.* 2015; VAHEDI *et al.* 2015). Importantly, functional studies abrogating these long-range interactions and the *trans* binding factors involved have provided evidence for spatiotemporal coordination of chromatin remodeling during these processes (AGARWAL *et al.* 1999).

Long-range interactions during skin development

More recently, long-range interactions have also been identified in several lineages of the skin, the most well characterized example being the Epidermal Differentiation Complex (EDC) located on human chromosome 1q21 that encompasses genes regulating cornified envelope formation. This locus is coordinated spatially and temporally, utilizing a conserved enhancer that establishes spatial interactions with these gene promoters to ensure exclusive expression within the differentiating keratinocytes of the stratifying epidermis (OH *et al.* 2014). Moreover, p63 is a transcriptional master regulator of epidermal differentiation that binds to a conserved enhancer to engage in looping interactions, and directly controls the expression of *trans*-binding factors (i.e. Satb1, Brg1) that are required for such higher-order chromatin configurations (ANTONINI *et al.* 2006; MARDARYEV *et al.* 2014). Recently, long-range interactions were reported at the *PPARγ2* and *IRF4* loci during adipocyte and melanocyte differentation, respectively (LEBLANC *et al.* 2014; VISSER *et al.* 2015). Collectively, these findings highlight the critical role of long-range interactions in the

spatiotemporal expression of key genes regulating skin development.

7. Work Described in This Thesis

The primary goal of my thesis research is to identify and characterize genetic mechanisms controlling human hair growth. To accomplish this, I have studied three inherited conditions affecting human hair growth as genetic models in which I performed detailed functional and molecular analyses of the causal genetic lesions and their downstream effects on gene expression in the hair follicle. The three major approaches to identify and characterize genetic mechanisms underlying human hair growth include: 1. Identifying genomic effects on human hair growth in rare, sporadic conditions (Chapter II), 2. Characterizing single-gene effects on human hair growth in rare, familial conditions (Chapter III), and 3. Functional analysis of rare, non-coding variants in a complex polygenic autoimmune disease, alopecia areata (Chapter IV).

The genetic architecture of rare, sporadic Mendelian disorders is usually characterized by large genomic rearrangements and copy number variants (CNVs) with large effect sizes. The Christiano lab and other groups have previously studied the genetic basis of inherited hypertrichoses, an extremely rare condition of excessive hair overgrowth that is often associated with additional congenital abnormalities (i.e. craniofacial, bone, and heart defects), identifying copy number variants and chromosomal rearrangements in autosomal and sporadic forms of the disease. In Chapter II of my thesis, I investigated the genetic mechanism associated with X-linked hypertrichosis (XLH) in a family in which the Christiano lab previously reported linkage to chromosome Xq24-27.

In collaboration with Drs. Brynn Levy and Dorothy Warburton, we performed cytogenetic analysis to detect structural variants underlying XLH. Using SOMA and FISH, we identified a 386 Kb duplication of chromosome 6 that had inserted onto the X chromosome. To identify the breakpoints and sequence through the insertion, we performed whole-genome sequencing and in collaboration with Dr. David Goldstein's laboratory whose members provided analytic tools and expertise, we identified a large interchromosomal insertion at chrXq27.1 that completely cosegregates with the phenotype. Our finding was consistent with previous reports of interchromosomal insertions in two other unrelated XLH families (ZHU *et al.* 2011). Remarkably, all three insertion events occurred at the same palindromic sequence at Xq27.1, yet the sequences contained within the insertions in each family are distinct. We then postulated that the presence of the insertion (rather than its content) may be responsible for the excessive hair overgrowth phenotype.

To determine the impact of the interchromosomal insertion on the expression levels of the surrounding genes, I performed RNA-seq and qRT-PCR that revealed a dramatic reduction in *FGF13* levels in affected individuals. Using several methods of detection, I found that *FGF13* is highly expressed in human hair follicles, and this expression is markedly reduced in XLH hair follicles. While we reported *FGF13* as the target of a position effect, it is highly likely that the interchromosomal insertion separates several genes from distal regulatory elements (i.e. enhancers) that are active in the hair follicle, and disrupts key intra- and/or interchromosomal interactions required for normal hair growth. Future studies using Chromosome Conformation Capture technologies will provide insight into perturbed or ectopic interactions in XLH.

In Chapter III, I studied a rare, familial condition of excessive hair overgrowth in a consanguineous family with autosomal recessive congenital generalized hypertrichosis terminalis (CGHT). The genetic architecture of rare, familial Mendelian disorders is usually composed of intragenic (exonic and splice site) mutations with large effect sizes. To identify rare, coding mutations in autosomal recessive CGHT, we performed whole-exome sequencing and using a standard filtering approach, identified a novel, rare splice variant in *ABCA5* that cosegregates with the phenotype in a homozygous recessive manner. Using immunohistochemistry and immunofluorescence staining, I found that *ABCA5* is highly expressed within the epithelial and mesenchymal compartments of both human and mouse skin and hair follicles. To determine the overall consequences of the mutation associated with CGHT, I then analyzed and quantified *ABCA5* mRNA and protein levels in CGHT patient cells, and visualized its localization pattern in affected skin and hair follicle tissues. Importantly, I found using several methods of detection that the *ABCA5* mutation not only leads to decreased protein levels in CGHT hair follicles, but also to a complete loss-of-function.

Since independent reports of Abca5 function in mouse tissues demonstrated a role for this ABC transporter in lysosome function and cholesterol efflux, I next tested whether CGHT patient cells possessed such defects at the cytological level. In collaboration with the Di Paolo laboratory at Columbia University, I analyzed and quantified lysosome function in the context of autophagy, and observed the free cholesterol distribution in patient keratinocytes. Importantly, cytological analyses in CGHT cells revealed defective autophagosome clearance and an overall accumulation of endolysosomal cholesterol, further supporting a loss-of-function mutation in *ABCA5*.

In parallel to these studies, we analyzed an unrelated sporadic case of CGHT using several cytogenetic techniques in collaboration with Drs. Brynn Levy, Dorothy Warburton, and Vaidehi Jobanputra. We identified a t3;17 translocation and cryptic 1.3 Mb deletion spanning *ABCA5,* creating an allelic series of six total CNVs at chr17q24.2-24.3 associated with CGHT. I hypothesized that *ABCA5* levels were disrupted in sporadic CGHT and although I did not identify an intragenic mutation on the opposite allele, I found using several methods of detection that both mRNA and protein levels were dramatically reduced in patient cells and hair follicles, suggesting haploinsufficiency of *ABCA5*. Importantly, these findings suggest a novel role for this transporter in regulating hair growth, and future functional studies using *Abca5* deficient mice will elucidate the precise mechanism(s) by which Abca5 regulates hair growth.

In Chapter IV, I investigated the genetic architecture of a complex, polygenic disease affecting hair growth by studying the role of rare, non-coding variants in alopecia areata (AA). AA is an autoimmune disease of hair loss resulting from the collapse of immune privilege in the hair follicle, and the Christiano lab previously performed the first Genome-Wide Association Study (GWAS) to determine the genetic architecture of AA, and identified the *ULBP3/6* locus on chr.6q25.1 (p = 5.9 x 10^{-24}) as the most significantly associated locus outside the *HLA* region. More recently, we conducted a meta-GWAS to expand gene discovery and identify true positive associations. We validated our previous findings and identified novel susceptibility loci (*ACOXL/BCL2L11(BIM)* on chr2q13 and *GARP (LRRC32)* on chr11q13.5) (Betz, Petukhova et al. 2015).

In Chapter IV.2, I characterized the expression of *BIM*, a gene that possesses known functions in apoptosis in melanocytes and the immune system (O'REILLY *et al.* 2000; LUO AND RUBINSZTEIN 2013). Moreover, I examined its expression pattern in mouse skin harvested in the growth (anagen), destructive (catagen), and resting (telogen) stages of the hair cycle. Interestingly, I observed BIM expression exclusively in catagen hair follicles, which are characterized by massive apoptosis. Using a mouse model that recapitulates the genetic and pathological basis of human AA, I found that BIM expression was no longer restricted to catagen hair follicles in lesional skin, but rather widespread and dramatically increased (Betz, Petukhova et al. 2015). Given the known role of BIM as a pro-apoptotic factor in the immune system, we postulate that its dysregulation contributes to the early entry of hair follicles into dystrophic catagen in active disease.

In Chapter IV.3, I investigated the molecular profiles of human and mouse AA. To identify dysregulated genes in AA, we performed transcriptional profiling using Affymetrix microarrays. I analyzed the expression data from the mouse model and identified key immune-related genes with dysregulated expression, which I validated using qRT-PCR and immunofluorescence staining in AA samples (XING *et al.* 2014). Using bioinformatics tools and unsupervised hierarchical clustering, we identified discrete gene expression signatures defined by key drivers of cytotoxicity, including IL15 and interferon gamma (IFNG), and observed a striking upregulation of these inflammatory cytokines and NKG2D ligands in both human and mouse lesional hair follicles. Using lymph node subcutaneous injections, we demonstrated that

CD8+NKG2D+ T cells are necessary and sufficient to induce disease in the mouse model, defining a role for the *ULBP*-NKG2D pathway in AA (XING *et al.* 2014).

In Chapter IV.4, I focused my studies on the *ULBP3/6* locus on chr.6q25 given the strong genetic association with AA ($P = 4.49 \times 10^{-19}$). Interestingly, this was the most significant region of association outside the *HLA,* and was not previously associated with any human disease. The *ULBP* genes encode ligands for the NKG2D cytotoxic T cell receptor, and we found in both human and mouse AA lesional hair follicles that NKG2D ligands are strongly upregulated (PETUKHOVA 2010; XING *et al.* 2014), further elucidating the involvement of the NKG2D/ULBP signaling axis in disease pathogenesis.

To identify human genetic variants at *ULBP3/6*, we performed targeted deep resequencing with Raindance technology and identified common and rare variants enriched on GWAS risk haplotypes. Using functional genomics, we uncovered a small number of rare, non-coding variants, three of which are novel and reside within *ULBP6* regulatory elements as well as CTCF binding sites. Since CTCF is an insulator protein involved in chromatin remodeling and gene repression, and *ULBP3/6* genes are not normally expressed in healthy hair follicles, I postulated that *ULBP6* non-coding variants disrupt transcription factor binding site motifs (i.e. CTCF) and/or regulatory elements required for *ULBP3/6* repression.

Using ChIP-PCR, I found that endogenous CTCF binding is enriched at the *ULBP6* candidate region in human scalp fibroblasts. Moreover, I demonstrated using *in vitro* reporter assays that the *ULBP6* region possesses regulatory activity that is reduced in the presence of AA rare variants, and that these variants abrogate CTCF-

mediated repression of the *ULBP6* sequence *in vitro*. Future studies examining this candidate region in long-range reporter assays and the long-range chromatin interactions *in vivo* will elucidate the mechanisms by which *ULBP3/6* gene expression is repressed in the healthy hair follicle, and whether CTCF-mediated interactions are disrupted in AA patient hair follicles.

Collectively, throughout my thesis, I used a wide range of methods and approaches to study three inherited conditions affecting hair growth, from rare, sporadic conditions to complex, polygenic diseases. Based on the inheritance pattern and disease prevalence of each condition, I employed different strategies to identify the underlying genetic variants, and these strategies were instrumental in my success identifying and characterizing novel genes and mechanisms controlling human hair growth.

Chapter II

Identifying genomic effects on hair growth

Structural variants and position effects in X-linked hypertrichosis

(Manuscript #1)

Chapter II.1 Preface

The genetic basis underlying rare, sporadic conditions is frequently characterized by structural rearrangements and/or copy number variants involving a locus or entire chromosome, rather than a single gene. While in some cases, the pathogenic variants can be detected using CNV analysis (i.e. by karyotyping), identifying complex rearrangements, microdeletions/duplications, and interchromosomal insertions has historically posed challenges, as these lesions were not as readily detectable using the conventional methods and approaches. In larger families with several affected members where the inheritance pattern is clear, positional information from linkage analysis has aided in narrowing the search. However, because these rare variants are sporadic and restricted to only the proband and affected family members, the power of the candidate approach is diminished and will most likely not detect the pathogenic defect.

Over the past decade, novel and more sophisticated technologies have been developed for genome sequencing, CNV analysis and SNP-based genotyping, and have proved useful in identifying the classes of mutations that were previously undetectable. Importantly, the completion of the Human Genome Project (2003) accelerated the development of these advanced technologies, thus facilitating mutational discovery. Importantly, an approach combining current methods for CNV analysis (i.e. SOMA), linkage (if applicable), and next-generation sequencing (i.e. whole-exome and whole-genome sequencing) is an optimal, robust and well-powered strategy to detect the rare and sporadic mutations in human disease. Upon the identification of candidate variants, a cosegregation analysis must be performed for families with more than one affected individual to establish whether there is a correlation

between the genotype and disease phenotype. To determine the involvement and mechanism(s) by which the candidate variants are causal to the pathology, further functional studies should be performed using *in vitro* (i.e. cell culture) and *in vivo* (i.e. mouse model) approaches.

In Chapter II.2, I investigated the genetic mechanism associated with X-linked hypertrichosis (XLH), a very rare condition of excessive hair overgrowth that the Christiano lab had previously studied over a decade ago. Linkage with microsatellite markers and haplotype analysis was performed in a large Mexican family with XLH segregating with deafness, palate and dental anomalies, which defined a 19 Mb minimal region on chrXq24-27 encompassing 82 genes. In collaboration with Dr. Conrad Gilliam, all 82 genes were sequenced using the Sanger method, but no mutations were found. Whole-exome sequencing was also performed, but this approach also failed to identify mutations in the candidate region. Several years following and in collaboration with Drs. Brynn Levy and Dorothy Warburton, we revisited this case to test our hypothesis that the underlying genetic lesion could be characterized by a structural/ numerical chromosomal abnormality rather than an intragenic mutation.

In Chapter II.2, we used SOMA to first detect CNVs or changes in dosage (such as insertions), followed by FISH to determine the location of the CNV or fragment. This approach was fruitful in identifying a copy number variant that created a structural variant via an interchromosomal insertion on the X chromosome (consistent with the inheritance pattern). Drs. Brynn Levy and Dorothy Warburton performed SOMA and analyzed the data. In order to identify the breakpoints as well as the content of the insertion, we performed whole-genome sequencing (WGS). We collaborated with Drs.

David Goldstein, Elizabeth Cirulli, and Yujun Han, who analyzed our WGS data using an in-house software program (SV-Finder) to sequence through the entire insertion. To determine whether all affected and carrier individuals contained this insertion, I then performed cosegregation analysis using a PCR-based approach to screen for the insertion breakpoints. Because skewed X-inactivation in female carriers is a hallmark feature of X-linked disorders, we tested X chromosome inactivation in XLH females. Drs. Maryam Shirazi and Vaidehi Jobanputra performed the X-chromosome inactivation assay and analyzed the data.

Interestingly, at the time we identified the genetic defect in this family, two other XLH families with linkage to chrXq24-27 were reported, where both families contained interchromosomal insertions on chrX. Remarkably, all three insertion events occurred at the same palindromic sequence at Xq27.1, however, the sequences contained within the insertions in each family are distinct. Therefore, we postulated that the presence of the insertion (rather than its content) might be responsible for the excessive hair overgrowth phenotype by disruption of the chromosomal architecture in the region. Consistent with our hypothesis, the sizes of the insertions (Kb) in each family and the severity of the phenotype with respect to the manifestation of additional congenital anomalies (i.e. deafness and craniofacial defects) greatly differs, inviting us to further investigate the downstream effects of the interchromosomal insertions.

In Chapter II.2, I investigated the mechanism(s) by which the interchromosomal insertion exerts pathogenic effects in XLH. Considering previous studies from the Christiano lab in another form of hypertrichosis (Ambras syndrome) that identified a position effect on the zinc-finger transcription factor, *TRPS1* as a result of CNVs and

chromosomal rearrangements, I hypothesized that the interchromosomal insertion in XLH might disrupt the expression levels of the neighboring genes. I then tested the expression levels of the neighboring genes on both sides of the insertion using qRT-PCR and RNA-seq on RNA isolated from whole skin, and Drs. Xiaoyun Sun and Yufeng Shen provided me with assistance in the RNA-seq data analysis. Drs. Julio Cesar Salas-Alanis and Rodrigo Cepeda obtained patient skin biopsies and clinical photos. Transcriptome sequencing revealed that among the genes surrounding the insertion, *FGF13* levels were significantly decreased in XLH.

FGF13 is an interesting candidate gene given the known role of FGFs in regulating hair growth (i.e. *FGF5* and the *angora* phenotype). I then determined whether *FGF13* was expressed in the human hair follicle, and using qRT-PCR, *in situ* hybridization, and immunofluorescence staining, I characterized its expression in both human scalp hair follicles and mouse hair follicles during morphogenesis and hair cycling. In the context of XLH, I tested *FGF13* levels using qRT-PCR and immunofluorescence analysis on keratinocytes and dermal fibroblasts, and patient hair follicles, respectively. Although in this study, we identified decreased levels of *FGF13* associated with XLH, it is more likely that larger effects involving other genes (including disrupted inter- and intrachromosomal interactions as a result of the insertion), may also contribute to the excessive hair overgrowth phenotype.

In Chapter II.3, I discuss these findings in a broader context and describe ongoing and future experiments to address the role of the interchromosomal insertion in the XLH pathogenesis.

Position effect on *FGF13* associated with X-linked congenital generalized hypertrichosis

Gina M. DeStefano[a], Katherine A. Fantauzzo[a,b], Lynn Petukhova[b,c], Mazen Kurban[b], Marija Tadin-Strapps[a], Brynn Levy[d], Dorothy Warburton[a,e], Elizabeth T. Cirulli[f], Yujun Han[f], Xiaoyun Sun[g], Yufeng Shen[g], Maryam Shirazi[d], Vaidehi Jobanputra[d], Rodrigo Cepeda-Valdes[h], Julio Cesar Salas-Alanis[h,i], and Angela M. Christiano[a,b,1]

Departments of [a]Genetics and Development, [b]Dermatology, [c]Epidemiology, and [d]Pathology and Cell Biology, Columbia University, New York, NY; [e]Department of Pediatrics, Columbia University Medical Center, New York, NY 10032; [f]Center for Human Genome Variation, Duke University School of Medicine, Durham, NC 27708; [g]Department of Biomedical Informatics, Columbia Initiative in Systems Biology, Columbia University, New York, NY 10032; [h]Dystrophic Epidermolysis Bullosa Research Association (DebRA), Nuevo Leon 67150, Mexico; and [i]Basic Science, Universidad de Monterrey, Nueva Leon 63238, Mexico

Edited by Elaine Fuchs, The Rockefeller University, New York, NY, and approved March 12, 2013 (received for review October 17, 2012)

X-linked congenital generalized hypertrichosis (Online Mendelian Inheritance in Man 307150) is an extremely rare condition of hair overgrowth on different body sites. We previously reported linkage in a large Mexican family with X-linked congenital generalized hypertrichosis cosegregating with deafness and with dental and palate anomalies to Xq24-27. Using SNP oligonucleotide microarray analysis and whole-genome sequencing, we identified a 389-kb interchromosomal insertion at an extragenic palindrome site at Xq27.1 that completely cosegregates with the disease. Among the genes surrounding the insertion, we found that Fibroblast Growth Factor 13 (*FGF13*) mRNA levels were significantly reduced in affected individuals, and immunofluorescence staining revealed a striking decrease in FGF13 localization throughout the outer root sheath of affected hair follicles. Taken together, our findings suggest a role for FGF13 in hair follicle growth and in the hair cycle.

congenital hypertrichosis | excessive hair growth

Inherited hypertrichoses are rare human disorders characterized by excessive hair growth that does not depend on androgen stimulation and is independent of age, sex, and ethnicity (1). Hypertrichosis syndromes fall under the larger umbrella of ectodermal dysplasias, or abnormal development of the hair, skin, nails, teeth, and/or eccrine glands, and are often associated with additional anomalies including gingival hyperplasia, deafness, cardiomegaly, and bone abnormalities (2). It has been suggested that inherited hypertrichoses represent examples of atavisms, or the recurrence of an ancestral phenotype, where the genes that promote a full coat of hair in other mammals and were silenced throughout evolution have become "reactivated" in human hypertrichosis, invoking unusual genetic mechanisms to explain their occurrence (3, 4).

We and others have identified genetic defects in two forms of autosomal dominant congenital generalized hypertrichosis (CGH) associated with copy number variants on chromosome 17q24 and with rearrangements on chromosomes 3, 7, and 8 (5–7). We previously reported a position effect on the zinc-finger transcription factor *Trichorhinophalangeal syndrome 1* (*TRPS1*) associated with Ambras syndrome congenital hypertrichosis, which was recapitulated in the koala (*Koa*) mouse hypertrichosis model (8). Likewise, we recently demonstrated a position effect on SRY-related HMG box gene 9 (*SOX9*) associated with CGH terminalis (9). Importantly, our findings provided evidence for position effects—instances in which a change in gene expression results from altering the location of a gene relative to its native chromosomal position (10)—as mechanisms contributing to inherited hypertrichoses (8).

In this work, we sought to identify the genetic mechanism associated with X-linked CGH in a family in which we previously reported linkage to chromosome Xq24-27 (11). We identified a large interchromosomal insertion that leads to decreased expression of a distant gene, Fibroblast Growth Factor 13 (*FGF13*), which is expressed in the hair follicle, providing evidence to support a position effect as the underlying genetic basis of X-linked hypertrichosis.

Results and Discussion

We ascertained a large kindred from Mexico with X-linked CGH (Online Mendelian Inheritance in Man 307150) cosegregating with deafness and with dental and palate anomalies (Fig. 1 *A–E*) (11). Affected males have approximately three times the number of normal hairs on the scalp and exhibit excessive growth of highly pigmented terminal hairs (medullated) on the scalp, back, shoulders, chest, arms, legs as well as on the face (Fig. 1 *A–D*), whereas hemizygous carrier females have mild hypertrichosis uniformly distributed across the body (Fig. 1*E*).

Histological analysis of affected hair follicles using hematoxylin and eosin staining confirmed that the hairs are of the terminal type because they are medullated, pigmented, and penetrate deep within the dermis (Fig. 1 *H* and *J*). Affected individuals have an increased density in the number of hair follicles and a transformation from vellus (fine, nonmedullated, unpigmented) to terminal hair follicles on multiple body sites, which cause an excessive hair overgrowth phenotype. Morphometric analysis of affected hair follicles revealed a widened dermal papilla (threefold increase; $P = 0.0000343$), matrix (1.9-fold increase; $P = 0.0000642$), and hair shaft (1.25-fold increase; $P = 0.036$) in hair follicles from three affected individuals compared with controls (Fig. 1 *G–J* and Fig. S1).

In our previous work on this family, we performed linkage analysis, which defined a 19-Mb region on Xq24-27 that spans ~82 genes and completely cosegregates with the disease (11). Sequencing the coding exons of every gene in the interval proved unsuccessful in identifying the mutation. To identify the genetic defect, we next performed SNP oligonucleotide microarray analysis (SOMA) using the Affymetrix Cytogenetics Whole-Genome 2.7M array, which revealed a 386-kb duplication of chromosome 6p21.2 (Fig. 2*A*). We then visualized the duplication at the cytogenetic level using fluorescence in situ hybridization (FISH) with two nonoverlapping BAC probes spanning the chromosome 6p21.2 duplication, which revealed a third signal for chromosome 6 present on the X chromosome (Fig. 2 *B* and *C*). To identify the insertion breakpoints as well as their content, we performed whole-genome sequencing (WGS), which revealed a large interchromosomal

Author contributions: A.M.C. designed research; G.M.D., K.A.F., L.P., M.K., M.T.-S., B.L., D.W., M.S., V.J., R.C.-V., and J.C.S.-A. performed research; E.T.C. and Y.H. contributed new reagents/analytic tools; G.M.D., K.A.F., L.P., E.T.C., Y.H., X.S., and Y.S. analyzed data; B.L. performed and analyzed FISH and SNP oligonucleotide microarray analysis data; D.W. performed and analyzed SNP oligonucleotide microarray analysis data; E.T.C. and Y.H. analyzed the whole-genome sequencing data; X.S. and Y.S. analyzed RNAseq data; M.S. and V.J. performed X-inactivation experiment; R.C.-V. and J.C.S.-A. obtained patient biopsies and clinical photos; and G.M.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: amc65@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216412110/-/DCSupplemental.

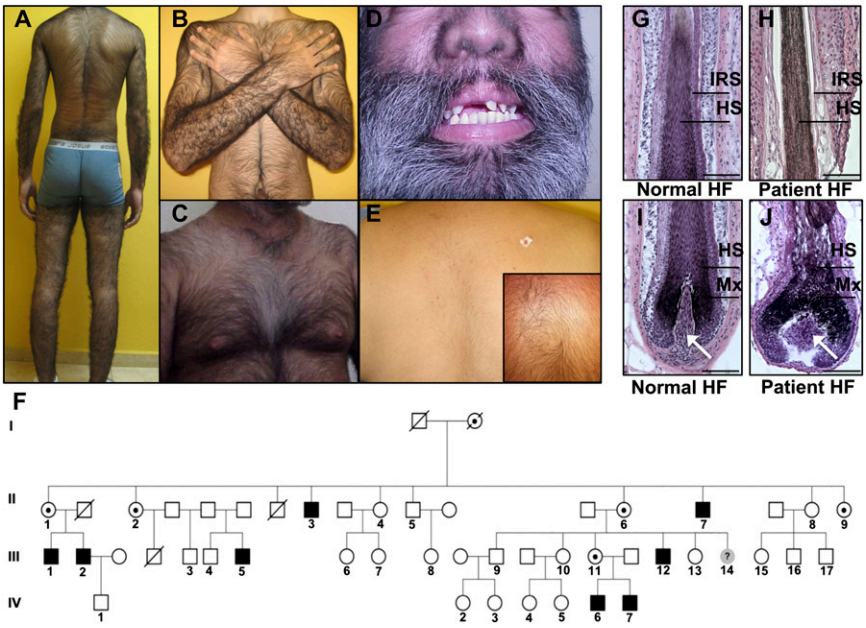


Fig. 1. Clinical features, pedigree, and histology of hair follicles in a Mexican family with X-linked CGH, deafness, and palate and dental anomalies. (*A–D*) Clinical photos of affected males with excessive hair growth on the back, shoulders, arms (*A* and *B*), chest (*C*), and (*D*) face. (*D*) Dental anomalies and a bulbous nose are also evident on patients. (*E*) Moderate hair growth is observed on the back of a female carrier. (*Inset*) A close-up image of a cowlick on the back of another carrier. (*F*) Pedigree of a four-generation family, three of whom are obligate carriers and eight of whom are affected. (*G–J*) Histology of a normal hair follicle and affected hair follicle, both from males, revealed that the hairs are of the terminal type, as they are medullated and highly pigmented. Affected hair follicles have a widened dermal papilla (white arrows), matrix (Mx), and hair shaft (HS) compared to control hair follicles. (Scale bars, 100 μm.)

insertion at an extragenic palindromic sequence on chromosome Xq27.1, consisting of a 386-kb duplication of chromosome 6p21.2 and 56 bp of chromosome 3q21.1 in the reverse orientation, separated by 14 bp of unknown origin (Fig. 3 *A* and *B*). This complex insertion leads to a 2-bp deletion within the Xq27.1 palindromic site and contains sequences from two genes on chromosome 6p21.2 [Dishevelled associated activator of morphogenesis 2 (*DAAM2*) and Kinesin family member 6 (*KIF6*)] and one gene from chromosome 3q21.2 [Family with sequence similarity 162, member A (*FAM162A*)], none of which have reported functions in the skin or hair follicle. To confirm that the insertion cosegregated with the disease phenotype, we used PCR amplification of the centromeric and telomeric breakpoint junctions on genomic DNA from control, carrier, and affected individuals, which revealed junction bands present only in affected and carrier individuals, whereas DNA from controls produced an amplicon representative of an unaffected X chromosome (Fig. 3 *C* and *D*).

To gain insight into whether the insertion at Xq27.1 affects X-chromosome inactivation in female carriers, because skewed X inactivation is a phenomenon reported in several X-linked disorders (12, 13), we performed the X-chromosome inactivation (XCI) assay and did not observe significant skewing (≥80%) in four of five carriers (Table S1 and *SI Materials and Methods*).

Importantly, two other recently reported families with X-linked CGH contain linkage to chromosome Xq24-27, consistent with our findings in which one family contains a 300-kb interchromosomal insertion from chromosome 4q31.2 and the other family contains a 125-kb interchromosomal insertion from chromosome 5q35.3 (14) (Fig. S2). Interestingly, the insertion events in all three families occur at the same human-specific extragenic palindrome sequence at Xq27.1, suggesting that the presence of the insertion (rather than its content) may be responsible for the excessive hair overgrowth phenotype by disruption of the chromosomal architecture in the region.

Despite the identification of different insertions in these three cases, their impact on the expression of the surrounding genes has not been thoroughly investigated. Therefore, we analyzed the expression of several neighboring genes using quantitative RT-PCR (qRT-PCR) on RNA isolated from control, carrier, and affected skin biopsies. Unexpectedly, we found that *FGF13* levels were significantly reduced in affected individuals by approximately fourfold ($P = 0.0007$) relative to controls and observed a clear dosage effect when comparing levels between carrier and affected individuals (Fig. 4 *A* and *B*). To verify that the change in *FGF13* expression observed in affected skin biopsies was not due to differences in the number of hair follicle cells present, we normalized *FGF13* expression to that of Keratin 14 (*KRT14*), which marks the outer root sheath and basal layer of the epidermis, and found that *FGF13* levels were dramatically reduced (18.1-fold; $P = 0.00006$) in affected hair follicle cells (Fig. 4*C*).

Importantly, the mRNA levels of additional neighboring genes were not significantly changed, and *SOX3* and Chromosome X open reading frame 66 (*Cxorf66*) expression levels were undetectable in control and affected individuals (Fig. 4*A*). To further examine the expression levels of the genes surrounding the 389-kb insertion, we performed RNA sequencing (RNA-seq) on RNA from the skin of one control and one affected individual, which verified the decrease in *FGF13* expression (eightfold; $P = 0.012$) and revealed that the expression levels of most of the genes in the surrounding region over a distance of ~3 Mb on either side of the insertion were not significantly changed (Table S2 and *SI Materials and Methods*). The expression of the microRNA, miR-504, intronic to *FGF13,* was reduced by approximately 1.5-fold in the carrier and affected individuals by qRT-PCR ($P = 0.00831$) (Fig. 4*A*). *FGF13* is not a predicted target gene of miR-504, yet there are several predicted targets with known roles in hair follicle development and cycling whose expression levels were altered in X-linked hypertrichosis, detected by RNA-seq (Table S3 and *SI Materials and Methods*). Although these changes in gene expression may simply reflect a difference in the number of hair follicle cells present, it remains possible that the reduction of *FGF13*–miR-504 transcripts either directly or indirectly leads to increased expression of some of these downstream genes.

FGF13 is a plausible candidate gene to be the target of the position effect because its expression was previously detected in the hair follicle bulge, which is the stem cell compartment of the hair follicle (15, 16). However, it was unclear whether these were the only *FGF13*-expressing cells in the human hair follicle or whether expression was more widespread. Using in situ hybridization and immunofluorescence staining on hair follicles in the growth stage of the hair cycle, anagen, we detected expression of FGF13 in the outer root sheath within the middle and upper portions of the human hair follicle (Fig. 4 *D* and *E*). We also observed FGF13 expression in the trichilemma, or outer root sheath compartment of the club hair, during the resting stage of the hair cycle, telogen, by immunofluorescence staining (Fig. 4*F*). In anagen hairs, FGF13 localizes to the outer root sheath (ORS) where KRT14 is expressed and also localizes to the companion layer that separates the outer from the inner root sheath, as evidenced by overlapping expression with KRT75 (companion layer marker) (Fig. S3 *A* and *B*). However, we did not observe FGF13

GENETICS

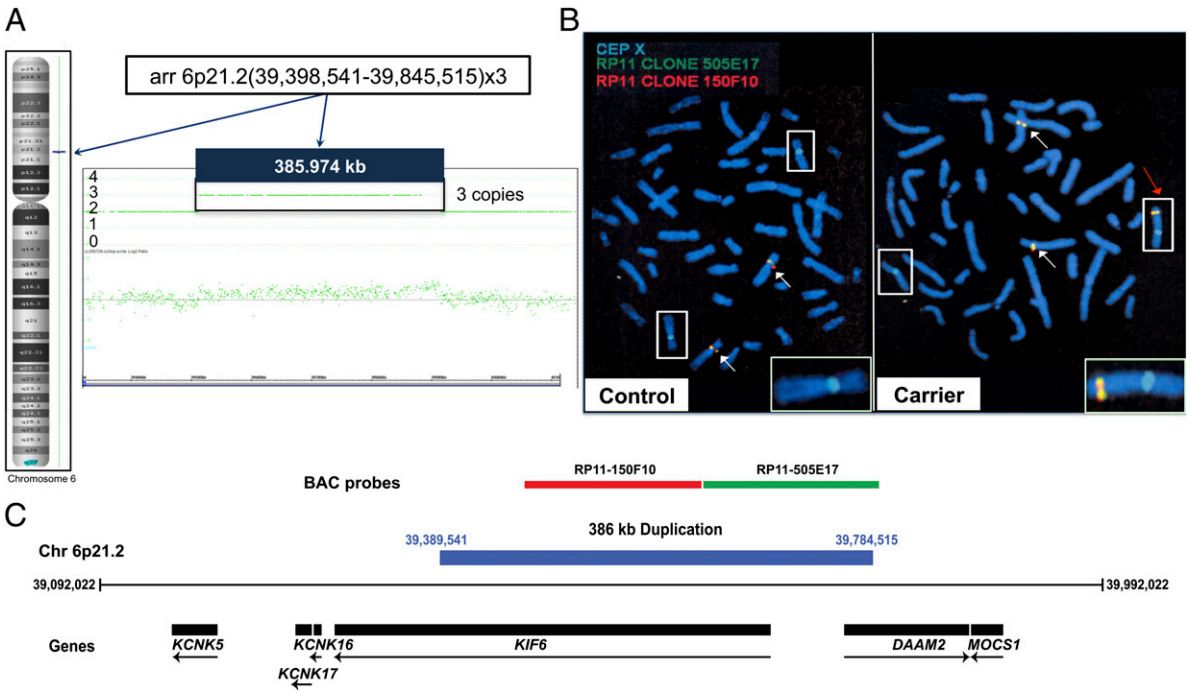


Fig. 2. SOMA identified a 386-kb duplication of chromosome 6, and FISH revealed its insertion on the X chromosome. (*A*) SOMA performed on an affected individual using the Affymetrix Cytogenetics Whole-Genome 2.7M array revealed a 386-kb duplication of chromosome 6p21.2 encompassing the *KIF6* and *DAAM2* genes (as shown in *C*). (*B*) FISH on control and carrier metaphase chromosomes revealed the insertion of the chromosome 6 duplication onto the X chromosome at the cytogenetic level. Boxes indicate X chromosomes, white arrows indicate chromosome 6, and the red arrow indicates the X chromosome containing the insertion. *Insets* are magnified images of the unaffected and affected X chromosome from control and carrier individuals, respectively. (*C*) Nonoverlapping BAC clones used to span the chromosome 6 duplication include the *KIF6* and *DAAM2* genes (drawn to scale). Genomic coordinates reference the UCSC Genome Browser human reference genome build hg19.

localization to the bulge region of the anagen human hair follicle, through costaining with CD200, a well-characterized bulge marker (Fig. S3*C*).

To gain insight into which cells exhibited decreased *FGF13* levels, we performed FGF13 immunofluorescence staining on control, carrier, and affected hair follicles and observed a decrease in the intensity of expression and the number of FGF13-positive cells throughout the outer root sheath of affected anagen follicles compared with controls and carriers (Fig. 5 *A* and *C*). Moreover, a comparison of affected and carrier telogen follicles revealed a decrease in expression throughout the affected hair follicle, recapitulating the dosage effect observed at the mRNA level (Fig. 5*B*).

To further investigate the selective decrease of *FGF13* expression in affected keratinocytes, we performed qRT-PCR on keratinocytes and fibroblasts cultured from control, carrier, and affected skin biopsies and observed a significant decrease in *FGF13* expression in affected keratinocytes of 6.7-fold (P = 0.00193) (Fig. 5*D*), but not in the fibroblasts. Consistent with our previous observations, our findings localize the defect to the keratinocyte compartment.

To determine *Fgf13* expression during murine hair follicle morphogenesis, we performed whole-mount and section in situ hybridization on embryonic day 12.5 (E12.5)–E16.5 embryos and observed strong expression in placodes (the sites of newly forming follicles), as well as in the dermal condensate beneath the placodes that becomes the dermal papilla of the hair follicle at E14.5 within the developing whisker pad and guard hair pelage follicles (Fig. S4*A*). Immunofluorescence staining of vibrissae follicles during morphogenesis at E16.5 revealed that Fgf13 localizes to the outer root sheath, similar to the postnatal localization pattern of the human FGF13 protein in anagen follicles (Fig. S4*B*). Moreover, immunofluorescence staining on postnatal skin revealed that Fgf13 localizes to the bulge, isthmus region, and outer root sheath of the hair follicle (Fig. S4 *C* and *D*). Taken together, our results suggest a potential role for Fgf13 in regulating hair follicle growth and cycling.

In mice and in humans, five-prime alternative splicing of *FGF13* and use of different transcription start sites generates transcripts with distinct 5′ exons referred to as 1S, 1U, 1V, 1Y, and 1V+1Y, where exons 2–5, encoding the conserved core region of the protein, are common to all transcripts. Isoform-specific PCR revealed that isoforms 1S, 1V, 1Y, and 1V+1Y are expressed in human scalp skin (Fig. S5). To gain insight into the mechanism by which the interchromosomal insertion alters *FGF13* transcript levels in X-linked hypertrichosis, we used the RNA-seq data to test for differentially expressed isoforms using Cuffdiff (*SI Materials and Methods*), but did not observe differential expression between the *FGF13* isoforms, suggesting that the interchromosomal insertion disrupts the transcription of all isoforms rather than altering the use of a particular transcription start site.

Position effects on single genes have been reported in several other human genetic diseases associated with large chromosomal rearrangements where the distances of the farthest breakpoints from the target genes were as large as 1.0 Mb [sonic hedgehog (*SHH*) in preaxial polydactylyl II] (17, 18), 1.3 Mb (*SOX9* in camptomelic dysplasia) (19, 20), and 7.3 Mb (*TRPS1* in Ambras syndrome) (8). Because *FGF13* lies 1.2 Mb away from the insertion and its expression was selectively reduced, we postulate that the interchromosomal insertion at Xq27.1 separates the gene from a tissue- or temporal-specific modifier element (such as an enhancer) required for proper *FGF13* expression during hair follicle morphogenesis and cycling. Consistent with this notion, we found that *FGF13* expression was selectively reduced in a tissue-specific manner, as transcript levels were decreased in affected keratinocytes, but not in fibroblasts.

Our data indicate that *FGF13* is primarily expressed throughout the outer root sheath of human hair follicles, and findings from the clinical, histological, and morphometric analyses revealed increased width of hair follicles in X-linked hypertrichosis (Fig. 1 *H* and *J* and Fig. S1). Several genes expressed in the outer root

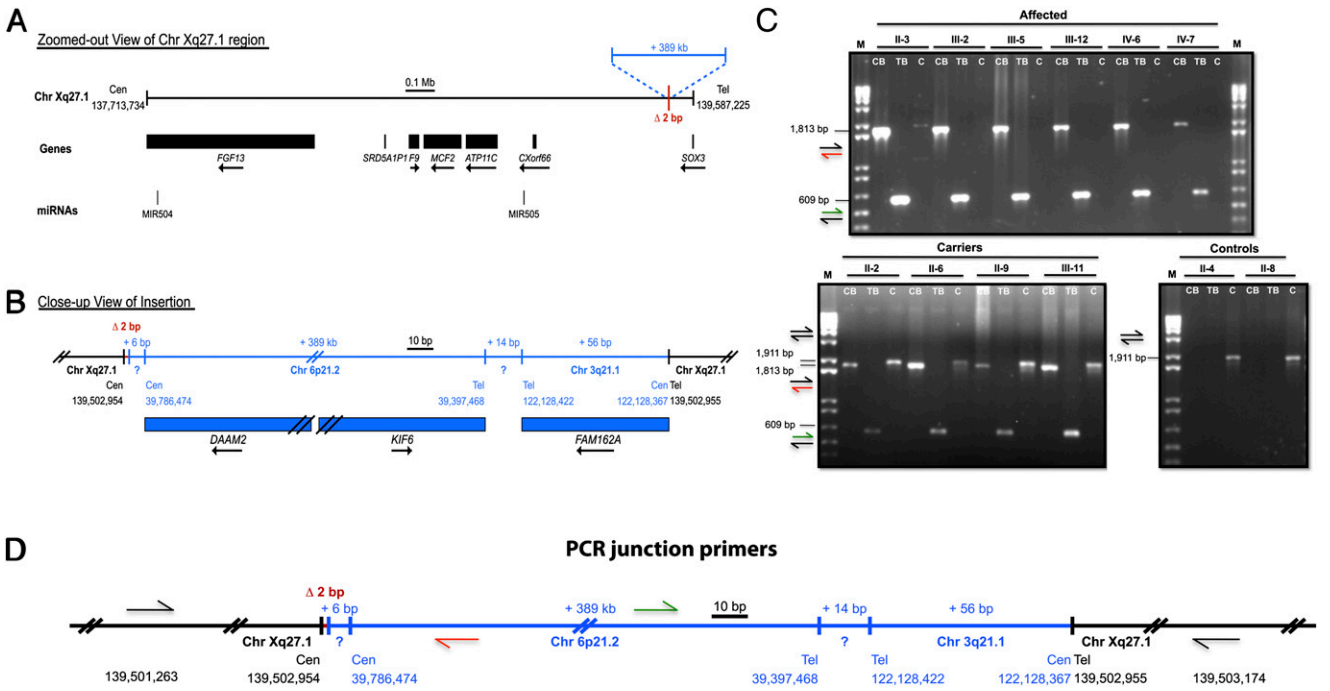


Fig. 3. Whole-genome sequencing revealed a 389-kb interchromosomal insertion at Xq27.1 that cosegregates with the X-linked hypertrichosis phenotype. (*A*) Chromosome Xq27.1 in X-linked hypertrichosis. The genes and microRNAs encoded in the surrounding region are shown as black boxes with arrows indicating the direction of transcription. (*B*) WGS was used to determine the breakpoints and content of the interchromosomal insertion (shown in blue), including the 386-kb duplication from chromosome 6p21.2, 14 bp of unknown origin, and 56 bp of chromosome 3q21.2. (*C*) PCR amplification of the centromeric and telomeric junctions of the insertion on DNA from control, carrier, and affected individuals demonstrated segregation of the X-linked phenotype in the family at the genomic level. CB, TB, and C represent centromeric breakpoint, telomeric breakpoint, and controls, respectively. M, marker (1 kb + ladder). (*D*) Primer design of the centromeric and telomeric junctions; colored arrows correspond to the amplicons produced as shown in *C*. All images are drawn to scale. Genomic coordinates reference the UCSC Genome Browser human reference genome build hg19.

sheath of hair follicles have been reported to have noncell autonomous effects on other compartments, indirectly or directly leading to changes in hair follicle width and/or length. In the case of *Dishevelled 2* (*Dvl2*), an effector of wingless-type MMTV integration site family, member 3 (Wnt3) signaling normally expressed in the outer root sheath and the precortical and precuticle cells of the hair shaft, its overexpression in the outer root sheath induces a short-hair phenotype by altering the differentiation of hair shaft precursor cells (21). Similarly, overexpression of *Vegf* in the outer root sheath, where it is normally expressed, induces perifollicular vascularization of the hair follicle, resulting in accelerated hair regrowth and in increased size of hair shafts (22).

In our mouse expression studies, we found that *Fgf13* is expressed during hair follicle induction and morphogenesis, suggesting that it may play an important role in these processes. Affected individuals in the X-linked hypertrichosis family possess an increased density of hairs; thus, it is possible that dysregulation of *FGF13* levels in X-linked hypertrichosis leads to the formation of extra hairs, but further studies using *Fgf13*-deficient mice would be needed to directly implicate a role for *Fgf13* in hair follicle morphogenesis and cycling. Interestingly, *Fgf13* expression has been demonstrated in the dental mesenchyme and developing tooth bud (23), an additional site of pathology for X-linked hypertrichosis patients who have dental and palate anomalies, suggesting a potential role for this gene in odontogenesis.

Several FGFs and their receptors are known to play important roles in hair regrowth. Although canonical FGFs are known to signal through their respective receptors to control hair growth via stem cell activation and quiescence, FGF13 is the first noncanonical FGF to be implicated in hair follicle morphogenesis and cycling. As FGF13 has been reported to bind the MAPK scaffolding protein islet brain 2 (IB2) (24), leading to activation of a stress-induced MAPK that lies downstream of the canonical FGF-signaling pathway (24), one possibility is that FGF13 internally modulates the transcriptional output of canonical FGF signaling to control hair growth. Consistent with this notion, we found the expression levels of several FGFs to be dysregulated in X-linked hypertrichosis (Table S4). Among these was *FGF5*, a known regulator of the anagen-to-catagen transition and responsible for the excessive hair overgrowth phenotype in angora mice, dogs, goats, and rabbits (25–28). A second possibility is that FGF13 acts as a microtubule-stabilizing protein in the hair follicle, similar to its role in neurons (29), to regulate additional signaling molecules active in the developing follicle. Further functional studies on FGF13 will reveal the mechanism by which it regulates hair follicle growth and distribution.

In this study, we identified a 389-kb interchromosomal insertion at Xq27.1 that completely cosegregates with the X-linked CGH phenotype and found that, among the genes surrounding the insertion, *FGF13* expression was selectively and profoundly reduced. *FGF13* lies 1.2 Mb away from the insertion, revealing a position effect on a distant gene as a result of the chromosomal insertion at Xq27.1. Although it has been suggested that these large interchromosomal insertions may mediate pathogenic effects by introducing novel regulatory elements, it is more likely that the presence of the insertions (rather than their content) is responsible for the hair overgrowth phenotype because the sequences contained within each insertion are different among the families (14). Moreover, these insertions occur at an extragenic palindrome sequence and do not disrupt the coding region of a gene in the surrounding region (Fig. S2).

The density of hair follicles covering the human body is markedly reduced compared with other primates, and as such, the excessive hair phenotype observed in hereditary hypertrichosis has been suggested to be reminiscent of an atavism (3). Various examples of atavisms involving several body parts

GENETICS

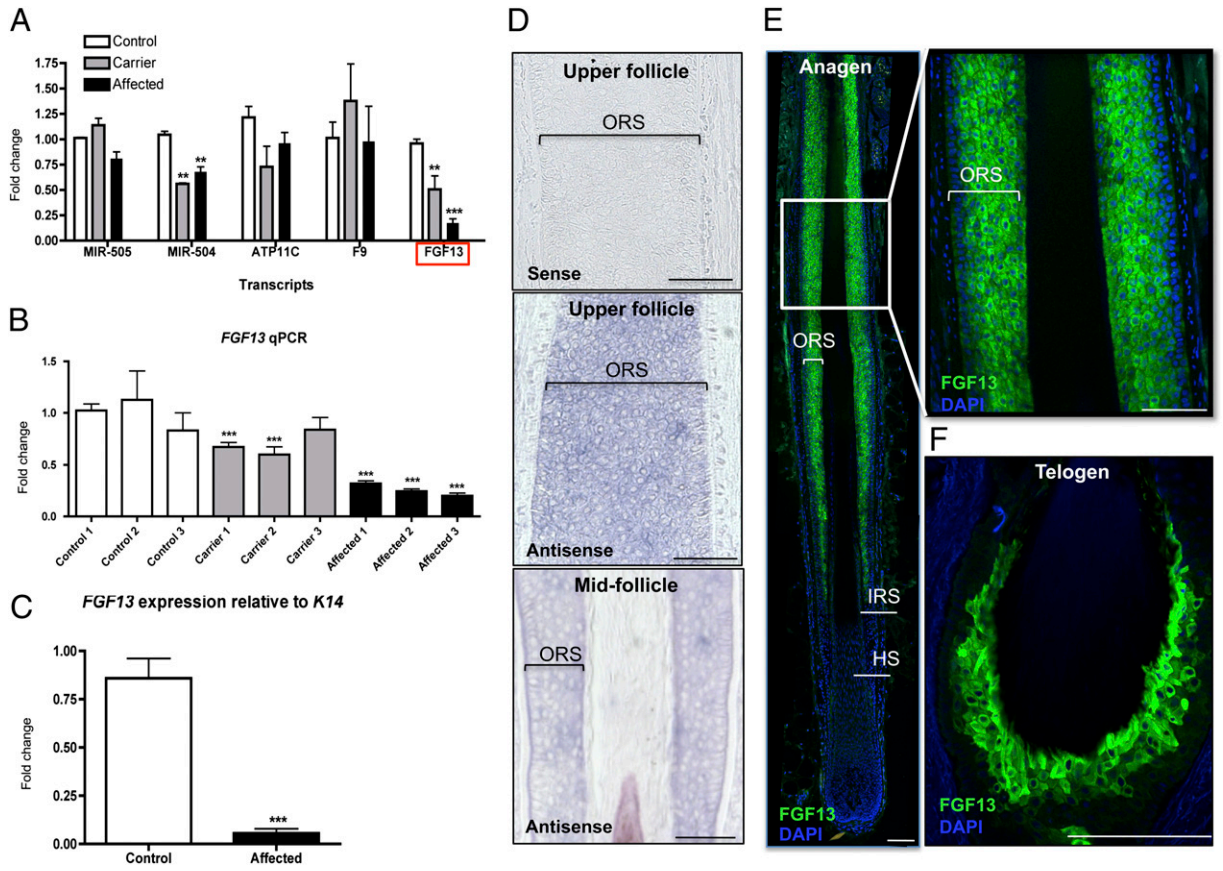


Fig. 4. *FGF13* levels are reduced in X-linked hypertrichosis, and FGF13 is expressed in the human hair follicle. (*A*) qRT-PCR of candidate genes surrounding the insertion on control, carrier, and affected skin biopsies reveals that *FGF13* levels are reduced by approximately fourfold in affected individuals relative to controls. (*B*) *FGF13* qRT-PCR in control, carrier, and affected individuals reveals a dosage effect. (*C*) *FGF13* expression normalized to *K14* reveals that *FGF13* levels are dramatically reduced by eighteen fold in affected cells relative to that of controls. qRT-PCR results are representative of the averaged values of three independent experiments on three controls, three carriers, and three affected individuals. Values are relative to the unaffected samples and standardized to the housekeeping gene *GAPDH* (in *A* and *B*). A Student *t* test was performed comparing each value to control 1 with a cutoff *P* value of 0.05 for statistical significance; $^{**}P < 0.01$; $^{***}P < 0.001$. Error bars represent the SEM. (*D*) In situ hybridization of *FGF13* in anagen hair follicles reveals expression in the outer root sheath (ORS) within the middle and upper portions of the hair follicle, where the sense probe produced no signal. (*E*) Immunofluorescence staining reveals that FGF13 localizes to the outer root sheath (magnified image) within the middle and upper portions of the human anagen hair follicle ($n = 5$). ORS, outer root sheath; IRS, inner root sheath; HS, hair shaft. (*F*) FGF13 expression is detected in the trichilemma (ORS) of telogen club-hair follicles by immunofluorescence staining. (Scale bar, 100 μm.)

have been reported in mammals, including the occurrence of complete ulnas and fibulas in miniature horses, reptile-like coronary circulation and myocardial architecture in humans, and the development of the ancestral tooth primordia in retinoic acid receptor-deficient mice (30–32). Importantly, the occurrence and prevalence of these ancestral features is highly suggestive of a genetic basis, one involving unusual mechanisms. Here, we report a position effect on *FGF13* in X-linked hypertrichosis that alters the spatiotemporal expression of the gene in the hair follicle. In light of our findings, we suggest that the altered *FGF13* expression in affected hair follicles influences important downstream signaling pathways, ultimately leading to the terminal hair overgrowth phenotype of X-linked hypertrichosis.

Materials and Methods

Ethics Statement. Informed consent was obtained from all subjects and approval for this study was provided by the Institutional Review Board of Columbia University in accordance with the Declaration of Helsinki Principles.

Patient Materials, Histological Analysis, and RNA Extraction. DNA was previously collected from 26 members of a family, 3 of whom are obligate carriers and 8 of whom are affected (11). Whole-skin biopsies taken from the back were then obtained from three female carriers and three affected male individuals. Biopsies were divided into three separate pieces for RNA extraction, cell culture [see *SI Materials and Methods* for details), and optimal cutting temperature compound (OCT) embedding for histological and morphometric analysis (see *SI Materials and Methods* for details]. We obtained control hair follicles from occipital scalp biopsies from discarded tissue following hair transplant surgeries. Samples were designated as non-human subject research under 45 Code of Federal Regulations (CFR) Part 46, and we therefore received an institutional review board exemption to use these materials. RNA extraction was performed using the Qiagen RNeasy Mini Kit following the manufacturer's instructions. Total RNA was used for first-strand cDNA synthesis, as previously described (8).

SOMA and WGS. DNA from an affected individual was prepared and hybridized as per the manufacturer's instructions on the Affymetrix Cytogenetics Whole-Genome 2.7M array, and data were analyzed with the Affymetrix Chromosome Analysis Suite. WGS was performed on one affected male of this Mexican family using the methods described in *SI Materials and Methods*. All coordinates reference the UCSC Genome Browser human reference genome build hg19.

Cytogenetic Analysis, Amplification of Genomic DNA, and qRT-PCR. FISH analysis was performed on metaphase chromosome spreads prepared from phytohemagglutinin (PHA)-stimulated cultured peripheral blood cells using standard techniques. The RPCI-11 clone 505E17 [labeled with Orange 5-TAMRA (carboxytetramethylrhodamine) dUTP] and RPCI-11 clone 150F10 (labeled with green 5-fluorescein dUTP) from Empire Genomics were used as FISH probes. Hybridization and posthybridization washing were performed as per the manufacturer's instructions. To test cosegregation of the insertion with

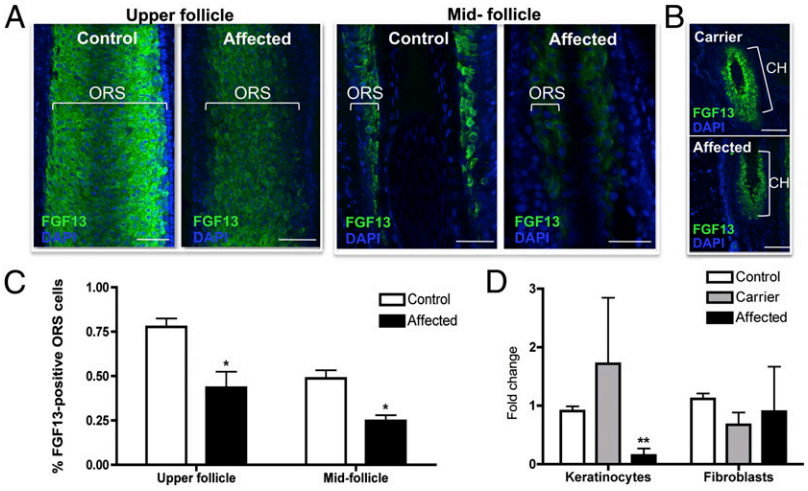


Fig. 5. Immunofluorescence staining reveals that FGF13 expression is dramatically reduced in affected hair follicles compared with control. (*A*) Immunofluorescence staining in control and affected anagen hair follicles reveals a decrease in FGF13 localization throughout the outer root sheath (ORS) in the mid and upper portions of the hair follicle. (*B*) Immunofluorescence staining in carrier and affected telogen hair follicles reveals decreased FGF13 expression in the affected hair follicle, recapitulating the dosage effect seen at the mRNA level. Z-stack images were taken using identical settings and a consistent Z-stack interval between control, carrier, and affected samples. CH, club hair of a telogen follicle. (*C*) Quantification of the percentage of FGF13-expressing ORS cells in control and affected hair follicles reveals a decrease in the number of FGF13-expressing cells within the upper and midfollicle regions of the ORS ($P < 0.05$). Data represent the averaged value of three independent experiments, where images taken at a 40× magnification were used to quantify the number of FGF13-positive cells relative to the total number of ORS cells. For immunofluorescence studies, hair follicles were stained from three control and two affected skin biopsies. (*D*) qRT-PCR revealed that *FGF13* expression is reduced in keratinocytes but not in fibroblasts grown from skin biopsies. A Student *t* test was performed with a cutoff *P* value of 0.05 for statistical significance; $*P < 0.05$, $**P < 0.01$. Error bars represent the SEM. (Scale bar, 100 μm.)

the disease phenotype, 100 ng of DNA was used for PCR amplification of the centromeric and telomeric breakpoints as well as of the control region of the unaffected X chromosome (details listed in the *SI Materials and Methods*). qRT-PCR was performed as previously described (8) using the Delta-Delta-Ct (ddCt) method, and primers are listed in the *SI Materials and Methods*.

In Situ Hybridization and Immunofluorescence Staining. Whole-mount in situ hybridization was performed as previously described (8) (see *SI Materials and Methods* for details). Immunofluorescence staining was performed on human control, carrier, and affected 12-μm hair follicle sections as well as on Swiss Webster dorsal skin sections (10 μm) from postnatal day 30 (anagen) and 50 (telogen) mice using the conditions described in the *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Drs. Antonio Sobrino, Jim Russo, Ming Chen, Barbara Ross, and Conrad Gilliam for their collaboration and stimulating discussions; Dr. Geoffrey Pitt for kindly providing the FGF13 antibody; Mr. Ming Zhang for technical assistance; and members of the A.M.C. laboratory for helpful discussions. We appreciate the support and expert assistance of the Skin Disease Research Center in the Department of Dermatology at Columbia University [National Institutes of Health (NIH)/National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Grant P30AR44535]. G.M.D. was supported by Columbia University Department of Genetics and Development Grant T32GM007088, and K.A.F. was supported by Columbia University Genetic Mechanisms of Skin Disease Grant T32AR007605. This work was supported in part by NIH/NIAMS Grant R01AR44924 (to A.M.C.).

1. Beighton P (1970) Congenital hypertrichosis lanuginosa. *Arch Dermatol* 101(6):669–672.
2. Garcia-Cruz D, Figuera LE, Cantu JM (2002) Inherited hypertrichoses. *Clin Genet* 61(5):321–329.
3. Cantú JM, Ruiz C (1985) On atavisms and atavistic genes. *Ann Genet* 28(3):141–142.
4. Hall BK (1984) Development mechanisms underlying the formation of atavisms. *Biol Rev Camb Philos Soc* 59(1):89–124.
5. Kim J, et al. (2010) Ambras syndrome in a Korean patient with balanced pericentric inversion (8)(p11.2q24.2). *J Dermatol Sci* 59(3):204–206.
6. Tadin M, et al. (2001) Complex cytogenetic rearrangement of chromosome 8q in a case of Ambras syndrome. *Am J Med Genet* 102(1):100–104.
7. Sun M, et al. (2009) Copy-number mutations on chromosome 17q24.2-q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia. *Am J Hum Genet* 84(6):807–813.
8. Fantauzzo KA, et al. (2008) A position effect on TRPS1 is associated with Ambras syndrome in humans and the Koala phenotype in mice. *Hum Mol Genet* 17(22):3539–3551.
9. Fantauzzo KA, Kurban M, Levy B, Christiano AM (2012) Trps1 and its target gene Sox9 regulate epithelial proliferation in the developing hair follicle and are associated with hypertrichosis. *PLoS Genet* 8(11):e1003002.
10. Kleinjan DJ, van Heyningen V (1998) Position effect in human genetic disease. *Hum Mol Genet* 7(10):1611–1618.
11. Tadin-Strapps M, et al. (2003) Congenital universal hypertrichosis with deafness and dental anomalies inherited as an X-linked trait. *Clin Genet* 63(5):418–422.
12. Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF (2002) Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am J Hum Genet* 71(1):168–173.
13. Knudsen GP, et al. (2006) Increased skewing of X chromosome inactivation in Rett syndrome patients and their mothers. *Eur J Hum Genet* 14(11):1189–1194.
14. Zhu H, et al. (2011) X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. *Am J Hum Genet* 88(6):819–826.
15. Ohyama M, et al. (2006) Characterization and isolation of stem cell-enriched human hair follicle bulge cells. *J Clin Invest* 116(1):249–260.
16. Kawano M, Suzuki S, Suzuki M, Oki J, Imamura T (2004) Bulge- and basal layer-specific expression of fibroblast growth factor-13 (FHF-2) in mouse skin. *J Invest Dermatol* 122(5):1084–1090.
17. Lettice LA, et al. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* 12(14):1725–1735.
18. Lettice LA, et al. (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci USA* 99(11):7548–7553.
19. Leipoldt M, et al. (2007) Two novel translocation breakpoints upstream of SOX9 define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. *Clin Genet* 71(1):67–75.
20. Velagaleti GV, et al. (2005) Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet* 76(4):652–662.
21. Millar SE, et al. (1999) WNT signaling in the control of hair growth and structure. *Dev Biol* 207(1):133–149.
22. Yano K, Brown LF, Detmar M (2001) Control of hair growth and follicle size by VEGF-mediated angiogenesis. *J Clin Invest* 107(4):409–417.
23. Kettunen P, Furmanek T, Chaulagain R, Kvinnsland IH, Luukko K (2011) Developmentally regulated expression of intracellular Fgf11-13, hormone-like Fgf15 and canonical Fgf16, -17 and -20 mRNAs in the developing mouse molar tooth. *Acta Odontol Scand* 69(6):360–366.
24. Schoorlemmer J, Goldfarb M (2002) Fibroblast growth factor homologous factors and the islet brain-2 scaffold protein regulate activation of a stress-activated protein kinase. *J Biol Chem* 277(51):49111–49119.
25. Hébert JM, Rosenquist T, Götz J, Martin GR (1994) FGF5 as a regulator of the hair growth cycle: Evidence from targeted and spontaneous mutations. *Cell* 78(6):1017–1025.
26. Housley DJ, Venta PJ (2006) The long and the short of it: Evidence that FGF5 is a major determinant of canine 'hair'-itability. *Anim Genet* 37(4):309–315.
27. Liu HY, Yang GQ, Zhang W, Zhu XP, Jia ZH (2009) [Effects of FGF5 gene on fibre traits on Inner Mongolian cashmere goats]. *Yi Chuan* 31(2):175–179.
28. Li CX, Jiang MS, Chen SY, Lai SJ (2008) [Correlation analysis between single nucleotide polymorphism of FGF5 gene and wool yield in rabbits]. *Yi Chuan* 30(7):893–899.
29. Wu QF, et al. (2012) Fibroblast growth factor 13 is a microtubule-stabilizing protein regulating neuronal polarization and migration. *Cell* 149(7):1549–1564.
30. Tyson R, Graham JP, Colahan PT, Berry CR (2004) Skeletal atavism in a miniature horse. *Vet Radiol Ultrasound* 45(4):315–317.
31. Walia I, Arora HS, Barker EA, Delgado RM III, Frazier OH (2010) Snake heart: A case of atavism in a human being. *Tex Heart Inst J* 37(6):687–690.
32. Peterkova R, Lesot H, Peterka M (2006) Phylogenetic memory of developing mammalian dentition. *J Exp Zoolog B Mol Dev Evol* 306(3):234–250.

Supporting Information

DeStefano et al. 10.1073/pnas.1216412110

SI Materials and Methods

Histological and Morphometric Analysis of X-linked hypertrichosis (XLH) Skin Biopsies. Whole-skin biopsies from the affected, carrier, and control individuals were embedded in optimal cutting temperature (OCT) compound, and a microtome cryostat was used to create 12-μm-thick hair follicle sections. Sections were stained with hematoxylin and eosin, permanently mounted with Permount (Thermo Fischer Scientific), and imaged using an HRc AxioCam fitted onto an Axioplan2 fluorescence microscope (Carl Zeiss). For morphometric analysis, the length-measuring tool in the AxioVision (release 4.8.2) program was used to calculate the distance between two points for each of the indicated hair follicle components; the widest distance for each structure was used, and the average value was taken using two to four measurements per section (with three to six sections per slide). Hair follicles were analyzed from one control and two affected individuals, where each skin biopsy contained two hair follicles, both of which were analyzed.

Isolation and Culture of Human Keratinocytes and Fibroblasts from Whole-Skin Biopsies. Keratinocytes and fibroblasts were grown from control, carrier, and affected skin biopsies using the following protocol: skin biopsies were collected in 10% (vol/vol) FBS in Dulbecco's Modified Eagle Medium (DMEM), washed with 5 mL PBS, and then chopped into small pieces that were transferred to 5 mL Dispase (5 mg/mL) overnight at 4 °C. Epidermis and dermis were separated with a scalpel, and the epidermis was placed into 5 mL 0.25% trypsin–EDTA at 37 °C for 30 min and then into 20 mL 10% (vol/vol) FBS in DMEM. Cells were collected by centrifugation at 1,000 × *g* for 7 min and resuspended in epidermal keratinocyte growth media, defined with supplements (CnT-07; CELLnTECH). Fibroblasts were isolated by digesting the dermis in 10 mL 0.3% collagenase for 4 h at 37 °C. Cells were collected by centrifugation at 1,200 × *g* for 10 min, washed in 30 mL fibroblast culture medium [10% (vol/vol) FBS in DMEM] twice, and then resuspended in fibroblast culture medium.

Whole-Genome Sequencing. DNA was prepared for sequencing according to the Illumina DNA sample preparation kit protocol. Initially, the DNA was randomly fragmented by nebulization followed by end repair, addition of a single A base, adaptor ligation, and gel electrophoresis to isolate 300-bp fragments followed by PCR amplification. Next, the size-selected libraries were used for cluster generation on the flow cell. All prepared flow cells were run on the Illumina HiSeq using the paired-end module: the paired-end reads were each 100 bp long. DNA was aligned to the reference genome (National Center for Biotechnology Information Build 36 Ensembl release 50) using the BWA software (version 0.4.9) (1). Picard was used to remove potential PCR duplicates via the rmdup command. SAMtools (version 0.1.5c) was used for variant identification, using the pileup command with the –c option and default settings (2). The variants were then filtered using a SAMtools variation filter with the default settings but removing the filter for a maximum allowed coverage per variant by setting it to 10 million. All variants were screened for quality by keeping only those with a consensus score and quality score of at least 20 [50 for insertions/deletions (INDELs)] and that had at least three reads supporting the variant. Heterozygous INDELs were also excluded if the ratio of variant reads to reference reads was less than 0.2. The average coverage for this sample was 44.4×. Large deletions and duplications were identified with the Estimation by Read Depth with SNVs software. Structural variants including insertions and translocations were identified using SV-Finder, software developed in the Duke Center for Human Genome Variation that uses multiple alignment-based approaches with an emphasis on split-read and pair-end.

The genomewide identification of functional gene variants was facilitated by SequenceVariantAnalyzer (3).

Amplification of Genomic DNA. The reaction conditions were as follows: 95 °C for 5 min, 94 °C for 30 s, 55 °C for 40 s, and 68 °C for 1.5 min, where 35 cycles were run with a final extension time of 10 min at 68 °C. The primers used for the control reaction were (F) TGGCATTACAAGAGTTAGCTTCTGA and (R) AATGCTTTGTAGTGGCTTTGTTTCC, producing an amplicon of 1,911 bp (4); the primers used for the centromeric breakpoint were (F) TGGCATTACAAGAGTTAGCTTCTGA and (R) CCTCCAGGGTGACTAAATTTG, producing an amplicon of 1,813 bp; and the primers used for the telomeric breakpoint were (F) AACTAGAAGGCCATTGGCTG and (R) AATGCTTTGTAGTGGCTTTGTTTCC, producing an amplicon of 609 bp.

Quantitative RT-PCR Analysis. The primers used in these assays were as follows: human Fibroblast Growth Factor 13 (*hFGF13*) (core)—(F) CAGCCGACAAGGCTACCAC and (R) GTTCCGAGGTGTACAAGTATCC; human MCF2 cell line derived transforming sequence (*hMCF2*)—(F) GCAGCAGGAACTTTTGACAG and (R) GCTGGTGTGTTCCAATTCAG; human SRY related HMG box 3 gene (*hSOX3*)—(F) GTTGGGACGCCTTGTTTAGC and (R) TAGCGCGAAGAAATATCAAACAG (4): human Coagulation Factor IX (*hF9*)—(F) GCATTCTGTGGAGGCTCTATC and (R) GCTGCATTGTAGTTGTGGTG; human ATPase, class VI, type IIC (*hATP11C*)—(F) GGACATTTCTGGCTGCCTTTG and (R) CCAGAATCGGGTATCCAAG; *hK14*—(F) GGGATCTTCCAGTGGGATCT and (R) GCAGTCATCCAGAGATGTGACC; and *hGAPDH*—(F) ATGGACACGCTCCCCTGACT and (R) GAAAGGTGGGAGCCTCAGTC. *hFGF13* isoform-specific PCR was performed using the following primers: *FGF13-001* (1S)—(F) CGAGAAATCCAACGCCTGC (5) and (R) CACCACTCGCAGACCCACAG; *FGF13-002* (1U)—(F) GTTAAGGAAGTCGTATTCAGAGC (5) and (R) CACCACTCGCAGACCCACAG; *FGF13-203* (1V)—(F) GATGCTTCTAAGGAGCCTCAG and (R) CACCACTCGCAGACCCACAG; *FGF13-202* (1Y)—(F) ACAGAGCCGGAAGAGCCTCAG and (R) CACCACTCGCAGACCCACAG; and *FGF13-201, -3* (1V+1Y)—(F) GATGCTTCTAAGGTTCTGGAT and (R) CACCACTCGCAGACCCACAG.

Expression was normalized to the GAPDH housekeeping gene and compared with the control samples. For each assay, cDNA was used from three controls, three carriers, and three affected individuals unless indicated otherwise. For expression analysis of microRNAs, human–miR-504 and human–miR-505, the following miScript primer assays (Qiagen) were used: Hs_miR-504_1, Hs_miR-505_1, and Hs_RNU6-2_1 miScript (miScript PCR Control). Images were generated using GraphPad Prism.

Whole-Mount and Section in Situ Hybridization. For mouse section in situ hybridization, dorsal skin isolated from Swiss Webster mice at indicated time points was harvested and embedded in OCT, where a microtome cryostat was used to generate 10-μm sections. For human studies, 12-μm hair follicle sections were used. The sense and antisense riboprobes were constructed using in vitro

transcription and the Digoxigenin (DIG)-labeling system (Roche). The following primers were used and recognize the core region of the *FGF13* sequence: mouse Fgf13 (*mFgf13*)—(F) TCAAACCAAGCTGTATTTGGC and (R) CTTTCAGTGGTTTGGGCAGAA; and *hFGF13*—(F) AGCCTCAGCTTAAGGGTATAG and (R) CAAGAACACTGTTACCTTGAGC.

Immunofluorescence Staining on Skin Biopsies. Sections were fixed with acetone at −20 °C for 10 min, washed three times in 1× PBS, and then blocked in 1.5% (vol/vol) fish gelatin/1% (vol/vol) BSA in 1× PBS at room temperature for 1 h. The rabbit anti-FGF13 antibody recognizing the C terminus of the protein was generously provided by Geoffrey Pitt (Duke University, Durham, NC) and was used at a concentration of 1:400 in 1.5% (vol/vol) fish gelatin/1% (vol/vol) BSA in 1× PBS. The rat anti-mouse CD200 antibody (1:100) (BD Pharmingen), guinea pig anti-human K75 (1:1,000) (a gift from Lutz Langbein, German Cancer Research Center, Heidelberg) and rabbit anti-human K14 (1:1,000) (Covance) were diluted in 1.5% (vol/vol) fish gelatin/1% (vol/vol) BSA in 1× PBS. The anti-rabbit, -rat, and -guinea pig IgG isotype (Santa Cruz Biotechnologies) antibodies were used as primary controls at the same concentrations as the respective primary antibodies listed above. Following PBS washes, the Alexa Fluor 488 donkey anti-rabbit IgG (Molecular Probes, Invitrogen), Alexa Fluor 594 donkey anti-rat IgG (Molecular Probes, Invitrogen), and Alexa Fluor 594 goat anti-guinea pig IgG (Molecular Probes, Invitrogen) secondary antibodies were added to the cryosections at a concentration of 1:800 in 1× PBS. Sections were mounted in VECTASHIELD mounting medium with DAPI (Vector Laboratories) and imaged using a LSM 5 laser-scanning Axio Observer Z1 confocal microscope (Carl Zeiss). For human studies, Z-stack images were taken at 10× and 20× magnifications using identical settings and consistent Z-stack intervals between slides. For mouse studies, images were taken at a 20× magnification.

Statistical Analysis. A Student *t* test (two-tailed) was used to determine statistical significance in quantitative RT-PCR assays with a significance level (α) of 0.05. Three biological replicates were used in each analysis (unless indicated otherwise), and values represent the average of three independent experiments for the three biological replicates. Error bars represent the SEM.

Assessment of X-Inactivation Skewing in Female Carriers. The human androgen receptor assay (HUMARA) was performed to determine skewing of X inactivation as previously described (6). Genomic DNA from five female carriers was used for amplification of a differentially methylated CpG dinucleotide site near a polymorphic region in exon 1 of the *Androgen Receptor* (*AR*) gene, and PCR products were digested with the HpaII (methylation-sensitive) and RsaI (cocutter, methylation-insensitive) restriction enzymes to distinguish between methylated and nonmethylated alleles. *X-chromosome-inactivation* skewing percentage was determined using the method described in ref. 6.

RNA Sequencing in Whole Skin. RNA sequencing (RNA-seq) was performed on whole skin from one control and one affected individual. Preparation of the cDNA library for sequencing was performed using the TruSeq kit (Illumina). In brief, 100 ng total RNA extracted from affected and control skin biopsies was purified (using polyA capture to select for mRNAs), fragmented, and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated and double-stranded cDNA was purified using bead capture. End repair was then performed to create blunt ends, followed by adenylation of the 3′ ends (to prevent intramolecular ligation), ligation of indexing adaptors to the ends of the double-stranded cDNA, and enrichment of DNA fragments containing adaptor molecules using PCR. The resulting cDNA library was then sent to the genomics core facility at The Rockefeller University to be sequenced on the Illumina HiSeq. 2000 machine using single-end reads of ∼50 bp, with an overall sequencing depth of ∼15 million reads per sample. Reads were mapped to the human reference genome (National Center for Biotechnology Information build 37.2) using TopHat, an algorithm designed to align reads from an RNA-Seq to a reference genome based on existing transcript annotation and inferred new splice sites (7). To estimate the relative abundance of genes and splice isoforms, the data were then analyzed using Cufflinks, a program that contains algorithms that estimate transcript abundance while accounting for alternative splicing (8). Fragments per kilobase of exon per million fragments mapped were normalized to the upper quartile. Differential expression of isoforms was tested using Cuffdiff, a program that uses the Cufflinks transcript quantification engine to determine transcript levels in more than one condition.

***In Silico* Prediction Analysis of miR-504 Target Genes.** miR-504 target genes were determined using a comprehensive database, miRWalk (9), which provides information on predicted, validated, and published microRNA target genes. We applied filters to select target genes with a minimum seed sequence of seven nucleotides, the longest transcript of a given gene, and a *P* value of 0.05 or less, which represents the strength of the prediction through a Poisson distribution. Furthermore, target genes that appeared in three or more of the following prediction programs were selected: TargetScan, miRanda, miRDB, PICTAR5, miRWalk, RNA22, and DIANA-mT.

1. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
2. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
3. Ge D, et al. (2011) SVA: Software for annotating and visualizing sequenced human genomes. *Bioinformatics* 27(14):1998–2000.
4. Zhu H, et al. (2011) X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. *Am J Hum Genet* 88(6):819–826.
5. Wang C, et al. (2011) Fibroblast growth factor homologous factor 13 regulates Na+ channels and conduction velocity in murine hearts. *Circ Res* 109(7):775–782.
6. Warburton D, et al. (2009) Skewed X chromosome inactivation and trisomic spontaneous abortion: No association. *Am J Hum Genet* 85(2):179–193.
7. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
8. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
9. Dweep H, Sticht C, Pandey P, Gretz N (2011) miRWalk–database: Prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform* 44(5):839–847.

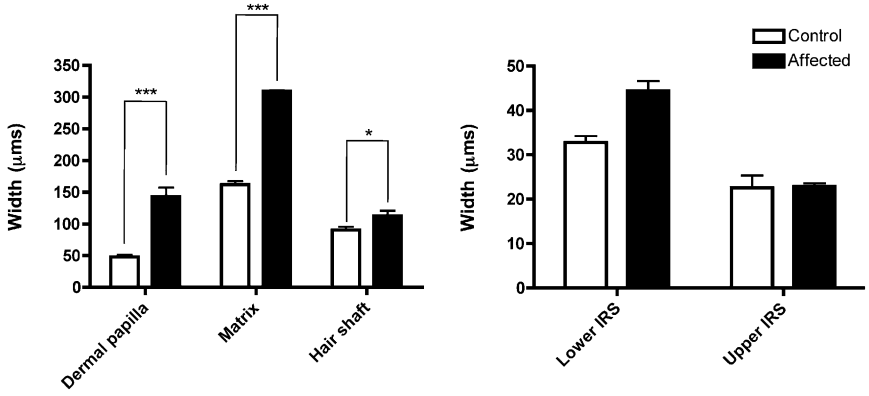


Fig. S1. Morphometric analysis of patient hair follicles reveals matrix, dermal papilla, and hair shaft defects. Quantification of hair follicle components using the length-measuring tool in the AxioVision program revealed a widened dermal papilla (threefold increase; $P = 0.0000343$), matrix (1.9-fold increase; $P = 0.0000642$), and hair shaft (1.25-fold increase; $P = 0.036$). Inner root sheath (IRS) differences in the lower and upper outer root sheath were not statistically significant. Hair follicles were analyzed from one control and two affected individuals, where the average widest distance was calculated for each hair follicle structure. A Student *t* test was performed comparing the affected to the control value with a cutoff *P* value of 0.05 for statistical significance; $*P < 0.05$; $***P < 0.001$. Error bars represent the SD.

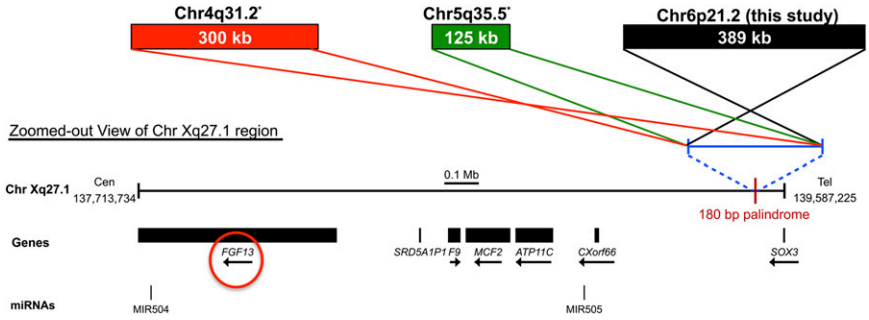


Fig. S2. Summary of interchromosomal insertion events in all three X-linked congenital generalized hypertrichosis families. Zoomed-out view of chromosome Xq27.1 region in X-linked hypertrichosis, where boxes indicate the insertion events and sizes of each from chromosomes 4, 5, and 6. Colored lines connected to the inverted triangle indicate orientation of the insertion events, where all three occur at the same palindromic sequence at Xq27.1. *, corresponds to ref. 1.

1. Zhu H, et al. (2011) X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. *Am J Hum Genet* 88(6):819–826.

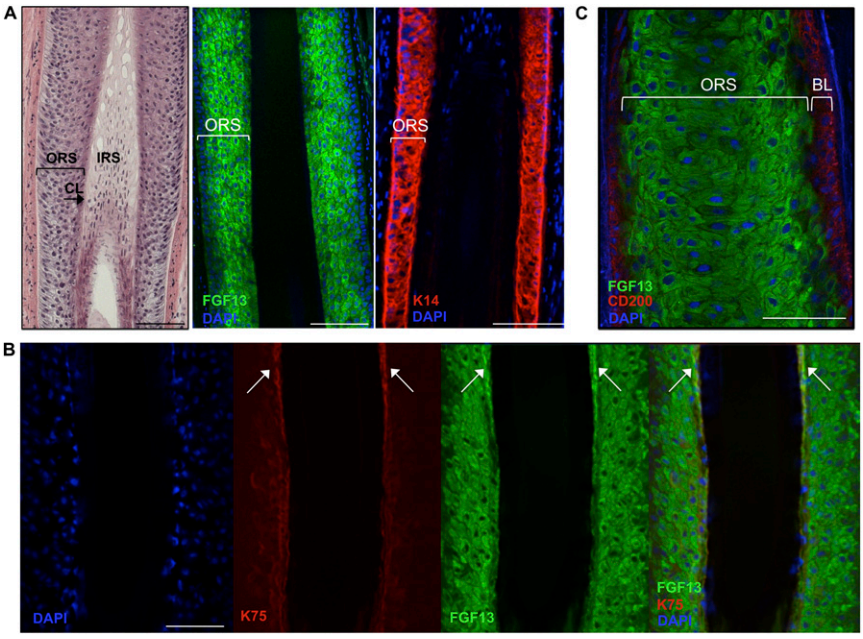


Fig. S3. FGF13 localizes to all layers of the outer root sheath (ORS) and the companion layer but not the human hair follicle bulge. (*A*) Immunofluorescence staining of FGF13 juxtaposed with Keratin 14 (KRT14) (which marks all layers of the ORS) demonstrates that FGF13 is broadly expressed throughout the ORS. The far right image is a hematoxylin-and-eosin staining of an anagen hair follicle for reference of morphology. (*B*) Costaining of FGF13 with KRT75, a marker of the companion layer between the ORS and inner root sheath (IRS), demonstrates that FGF13 localizes to the companion layer (arrows). (*C*) Costaining of FGF13 and CD200 (a marker of the bulge region of the hair follicle) demonstrates that FGF13 does not localize to the human hair follicle bulge. BL, basal layer. (Scale bar, 100 μm.)

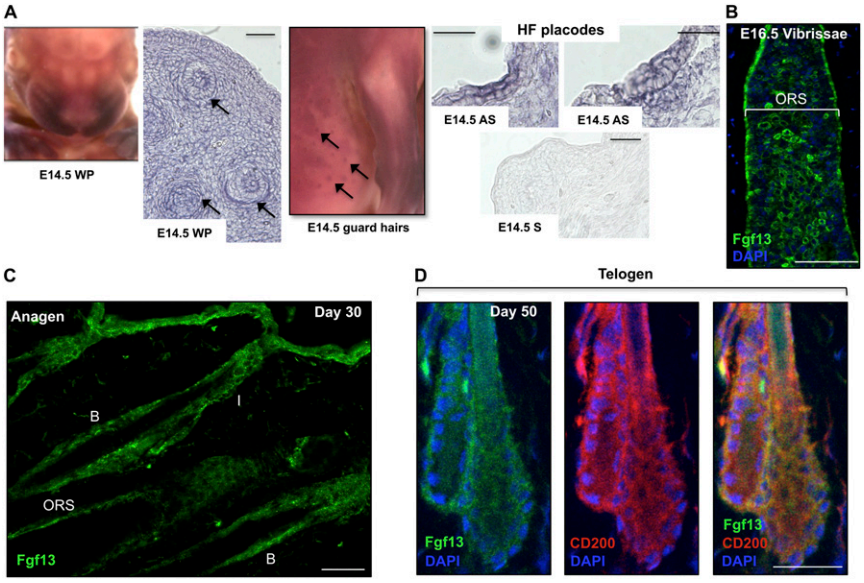


Fig. S4. *Fgf13* is expressed in the developing and cycling mouse hair follicle. (*A*) Whole-mount and section in situ hybridization of *Fgf13* on embryonic day 14.5 (E14.5) embryos reveals expression in the developing whisker pad (WP), guard hair follicle placodes, and dermal condensates at E14.5 ($n = 3$; AS, antisense; S, sense). Arrows indicate vibrissa and guard hair follicles. (*B*) Immunofluorescence staining on E16.5 vibrissae follicles demonstrates that Fgf13 localizes to the outer root sheath (ORS) ($n = 3$). (*C*) Immunofluorescence staining of Fgf13 in anagen (day 30) hair follicles reveals that Fgf13 localizes to the bulge (B), isthmus (I), and ORS. (*D*) Costaining of Fgf13 and CD200 in telogen (day 50) hair follicles reveals that Fgf13 localizes to the bulge. (Scale bar, 100 μm.)

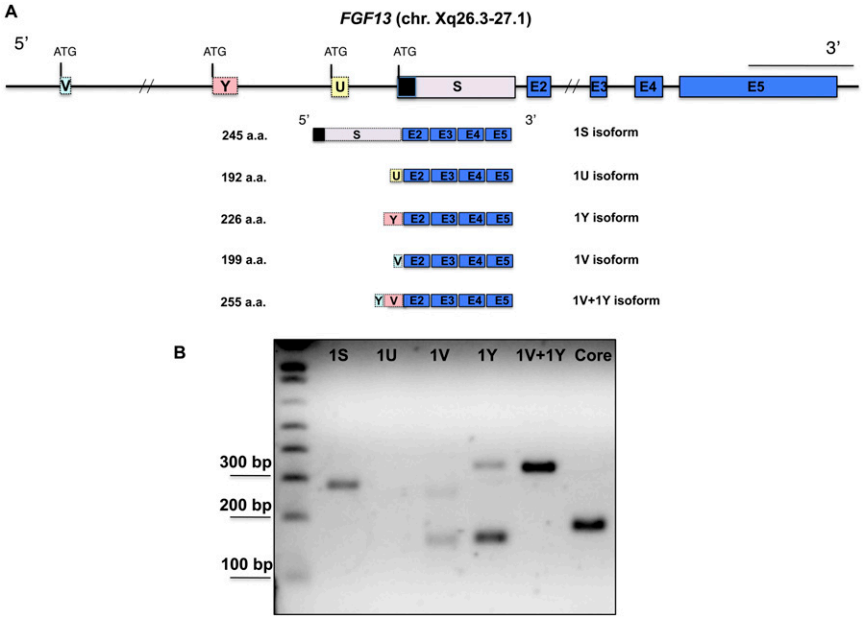


Fig. S5. Isoform-specific PCR of *FGF13* in whole skin. (*A*) Schematic of the *FGF13* locus at chromosome Xq26.3–27.1 (chr. Xq26.3–27.1). Alternating 5′ exons (termed 1S, 1U, 1V, 1Y, and 1V+1Y) are represented as boxes with distinct colors, whereas exons 2–5, common to all transcripts, are shown as blue boxes. The dark box at the 5′ end of the 1S isoform represents a nuclear localization signal. (Scale bar at top right, 100 kb.) (*B*) Amplification of *FGF13* transcripts using cDNA from whole skin demonstrates that the 1S, 1Y, and 1V+1Y isoforms are strongly expressed, and the V isoform is faintly expressed. "Core" represents the 3′ region common to all these isoforms. The Ensembl transcripts corresponding to these splice variants are *FGF13-001* (1S), *FGF13-002* (1U), *FGF13-203* (1V), *FGF13-202* (1Y), and *FGF13-201, -3* (1V+Y).

Table S1. X chromosome inactivation experiment in XLH female carriers

Sample ID	Description	% skewing
II-2	Female carrier	43
II-9	Female carrier	81
II-1	Female carrier	67
II-6	Female carrier	62
III-11	Female carrier	74
II-8	Female nonaffected	Noninformative
III-8	Female nonaffected	Noninformative

Table S2. Differential expression of FGF13 (bolded) and other genes ~3 Mb on either side of the 389-kb insertion (underlined and italicized) at Xq27.1 by RNA-seq

Gene name	Locus	Detectable expression	ln(fold change affected vs. unaffected)	Fold change (affected vs. unaffected)	*P* value
ZIC3	Chr:136476011–136481925	No	N.D.	N.D.	0.213
LOC158696	Chr:137524557–137527465	No	N.D.	N.D.	0.159
FGF13	**Chr:137541399–138114851**	**Yes**	**2.076**	**−7.976**	**0.012**
MIR504	Chr:137541399–138114851	No	N.D.	N.D.	1.000
LOC100129662	Chr:137541399–138114851	No	N.D.	N.D.	1.87*E*-05
Pseudogene (SRD5A1P1)	Chr:138356643–138358798	Yes	1.562	−4.770	0.005
F9	Chr:138440560–138473283	No	N.D.	N.D.	0.023
MCF2	Chr:138491595–138618047	Yes	0.320	−1.376	0.501
ATP11C	Chr:138636170–138742113	Yes	0.184	−1.201	0.663
MIR505	Chr:138833972–138834056	No	N.D.	N.D.	1.000
CXorf66	Chr:138865549–138875343	No	N.D.	N.D.	1.000
Insertion at chrX:139,330,620					
SOX3	Chr:139412817–139414891	No	N.D.	N.D.	1.000
RP1-177G6.2	Chr:139619589–139624662	No	N.D.	N.D.	0.677
CDR1	Chr:139693090–139694389	Yes	0.567	−1.763	0.387
MIR320D2	Chr:139836002–139836050	No	N.D.	N.D.	1.000
SPANXB1	Chr:139924426–139925542	No	N.D.	N.D.	1.000
LDOC1	Chr:140097596–140098976	Yes	0.382	−1.465	0.484
SPANXC	Chr:140163261–140164312	No	N.D.	N.D.	1.000
SPANXA1	Chr:140418508–140565735	No	N.D.	N.D.	1.000
SPANXA2	Chr:140418508–140565735	No	N.D.	N.D.	1.000
SPANXA2-OT1	Chr:140418508–140565735	No	N.D.	N.D.	0.079
SPANXD	Chr:140613233–140614321	No	N.D.	N.D.	1.000
SPANXE	Chr:140613233–140614321	No	N.D.	N.D.	1.000
MAGEC3	Chr:140753767–140813284	No	N.D.	N.D.	0.159
MAGEC1	Chr:140819307–140824853	No	N.D.	N.D.	1.000
MAGEC2	Chr:141117793–141120742	No	N.D.	N.D.	1.000
SPANXN4	Chr:141941369–141949732	No	N.D.	N.D.	1.000

Genomic coordinates reference UCSC Genome Browser human reference hg18. ln, natural logarithm; N.D., not detectable.

Table S3. miR-504 predicted target genes differentially expressed in XLH by RNA-seq

Gene name	Locus	Fold change (affected vs. unaffected)	ln(fold change)	*P* value	*Q* value
BATF2	Chr11:64511992–64521093	40.711	−3.707	0.001	0.008
FZD9	Chr7:72486044–72488386	34.025	−3.527	1.80*E*-05	0.000
TGM4	Chr3:44891101–44931092	31.450	−3.448	0.002	0.015
NPTX1	Chr17:76055227–76064999	30.048	−3.403	3.92*E*-09	2.91*E*-07
CRTAM	Chr11:122214464–122248557	26.236	−3.267	0.003	0.022
HTR7	Chr10:92490555–92607651	22.499	−3.113	0.000	0.002
UBASH3B	Chr11:122031607–122190397	21.982	−3.090	2.47*E*-08	1.49*E*-06
FUT5	Chr19:5816836–5821551	19.225	−2.956	2.66*E*-05	0.001
ATP8B3	Chr19:1733073–1763270	18.705	−2.929	2.83*E*-08	1.67*E*-06
MDGA1	Chr6:37708261–37773744	16.793	−2.821	6.99*E*-07	2.79*E*-05
FJX1	Chr11:35596310–35598997	11.675	−2.457	5.67*E*-06	0.000
KIF21B	Chr1:199205136–199259451	11.648	−2.455	6.60*E*-13	9.76*E*-11
MET	Chr7:116099694–116225676	10.666	−2.367	1.44*E*-06	5.17*E*-05
TP53RK	Chr20:44619868–44751683	10.417	−2.343	4.26*E*-05	0.001
MALL	Chr2:110198735–110231432	10.377	−2.340	1.09*E*-05	0.000
IL6R	Chr1:152644292–152708550	10.361	−2.338	1.09*E*-08	7.38*E*-07
STC1	Chr8:23755378–23768265	10.269	−2.329	1.39*E*-05	0.000
HAS3	Chr16:67696967–67723994	9.828	−2.285	1.78*E*-08	1.13*E*-06
S100A7A	Chr1:151655623–151662325	8.701	−2.163	3.66*E*-05	0.001
MTHFD2	Chr2:74279197–74295932	8.430	−2.132	2.00*E*-06	6.96*E*-05
CDCP1	Chr3:45098772–45162918	8.277	−2.113	5.62*E*-05	0.001
EXOSC6	Chr16:68841634–68843334	8.037	−2.084	0.000	0.002
TNFRSF9	Chr1:7898517–7925812	8.020	−2.082	0.004	0.030
P2RY11	Chr19:10077898–10091599	7.800	−2.054	0.000	0.002
LPCAT1	Chr5:1514541–1577076	7.264	−1.983	0.000	0.003

Table S3. Cont.

Gene name	Locus	Fold change (affected vs. unaffected)	ln(fold change)	*P* value	*Q* value
LIF	Chr22:28966441–28972796	6.932	−1.936	0.000	0.003
PNO1	Chr2:68238508–68256595	6.650	−1.895	0.000	0.005
BYSL	Chr6:41996942–42008762	5.907	−1.776	0.001	0.009
TRIB3	Chr20:309307–326203	5.706	−1.742	0.001	0.010
PCDH10	Chr4:134289919–134332182	5.591	−1.721	0.005	0.033
BCAT1	Chr12:24854224–24993660	5.357	−1.678	0.001	0.010
CNO	Chr4:6768742–6770288	5.247	−1.658	0.003	0.023
CYGB	Chr17:72035024–72053053	5.226	−1.654	0.005	0.035
ZFAT	Chr8:135559212–135794474	5.216	−1.652	8.57*E*-08	4.52*E*-06
STK40	Chr1:36577811–36624072	5.210	−1.651	0.002	0.016
SPRED2	Chr2:65391488–65513160	5.123	−1.634	9.36*E*-05	0.002
FUT2	Chr19:53891039–53901003	5.119	−1.633	0.000	0.005
IPPK	Chr9:94099460–94472368	5.105	−1.630	0.002	0.015
PXN	Chr12:119123476–119187957	5.053	−1.620	5.16*E*-08	2.90*E*-06
CD86	Chr3:123256898–123322678	5.052	−1.620	0.001	0.007
ABL2	Chr1:177335084–177465442	5.027	−1.615	9.74*E*-13	1.43*E*-10
KLHL21	Chr1:6573370–6585516	4.985	−1.606	0.007	0.044
FSCN1	Chr7:5598961–5612813	4.963	−1.602	0.004	0.031
RUNX1	Chr21:35081967–35343465	4.835	−1.576	2.14*E*-05	0.001
RRAGC	Chr1:39077605–39097927	4.730	−1.554	0.002	0.020
SERTAD1	Chr19:45620248–45623772	4.697	−1.547	0.003	0.022
OSBP2	Chr22:29420792–29633811	4.695	−1.547	0.005	0.033
ZNF264	Chr19:62394679–62426026	4.653	−1.538	0.004	0.028
DNMBP	Chr10:101625323–101759666	4.581	−1.522	0.003	0.023
PLEKHM1	Chr17:40869048–40923929	4.558	−1.517	2.80*E*-05	0.001
PDRG1	Chr20:29996418–30003544	4.414	−1.485	0.007	0.042
MBP	Chr18:72819776–72973762	4.300	−1.459	3.75*E*-05	0.001
PLCD3	Chr17:40544533–40565417	4.266	−1.451	0.005	0.033
SLC7A1	Chr13:28981550–29067825	4.162	−1.426	0.007	0.043
GMEB2	Chr20:61689398–61728825	4.160	−1.425	0.007	0.041
LRRC8A	Chr9:130684211–130720138	4.158	−1.425	0.000	0.006
KCNK1	Chr1:231816372–231874881	4.037	−1.396	0.006	0.039
NUDT15	Chr13:47509703–47519283	4.004	−1.387	0.008	0.046
LY6K	Chr8:143778530–143805393	3.958	−1.376	0.001	0.006
SMURF1	Chr7:98462993–98579679	3.939	−1.371	0.002	0.016
GPER	Chr7:1003148–1144419	3.871	−1.354	0.008	0.047
RRAS2	Chr11:14256041–14342628	3.637	−1.291	0.000	0.002
CRTC3	Chr15:88874201–88989581	3.559	−1.270	0.001	0.010
MAPKAPK2	Chr1:204924987–204974253	3.464	−1.242	0.005	0.035
ZBTB24	Chr6:109890411–109911133	3.342	−1.207	0.008	0.047
NRF1	Chr7:129038790–129184158	3.334	−1.204	0.006	0.038
DIAPH1	Chr5:140874771–140978806	3.312	−1.197	0.001	0.011
PAPD5	Chr16:48744329–48826720	3.160	−1.151	0.008	0.045
DHX33	Chr17:5284955–5313104	3.083	−1.126	0.003	0.022
PVR	Chr19:49838937–49861268	2.872	−1.055	0.001	0.009
TGFBR1	Chr9:100907232–100956294	2.854	−1.049	0.008	0.046
GAS7	Chr17:9754650–10042593	2.703	−0.994	0.006	0.036
IL1RAP	Chr3:191714533–191857680	2.632	−0.968	0.002	0.015
CTDP1	Chr18:75540788–75615498	2.594	−0.953	0.007	0.044
VEGFA	Chr6:43845923–43862201	2.490	−0.912	0.000	0.005
CNOT4	Chr7:134697086–134845415	2.220	−0.798	0.004	0.026

Genomic coordinates reference UCSC Genome Browser human reference hg18. ln, natural logarithm.

Table S4. FGFs with dysregulated expression levels in XLH by RNA-seq in whole skin

Gene name	Locus	Fold change (patient vs. control)	*P* value
FGF5	Chr4:81406765–81431195	−56.33	2.53*E*-09
FGF18	Chr5:170779271–170817235	−11.22	3.00*E*-05
FGF14	Chr13:101171205–101852125	−9.13	0.0004
FGF13	Chr:137541399–138114851	−7.98	0.012
FGF12	Chr3:193339875–193928082	−6.09	0.006
FGFR1OP	Chr6:167332805–167374056	−3.37	0.007
FGFR3	Chr4:1764836–1780397	−3.21	0.001
FGFR2	Chr10:123227833–123347962	−2.32	0.0008
FGFBP1	Chr4:15546289–15549461	13.58	3.70*E*-06

Genomic coordinates reference UCSC Genome Browser human reference hg18.

Chapter II.3

Discussion

Chapter II.3 Discussion

The density of hair follicles across the human body is regionally decreased on different body sites compared to other mammals. Thus, the extreme excessive hair overgrowth phenotype observed in inherited hypertrichosis has long fascinated evolutionary geneticists. Historically, it has been considered an atavism, or recurrence of an ancestral phenotype (HALL 1984; CANTU AND RUIZ 1985). The estimated incidence of these syndromes in the general population is extremely low, with approximately 60 reported cases to date, suggesting an unusual underlying genetic mechanism whose observed frequency is just as rare, and likely characterized by unusual, sporadic mutations with pleiotropic effects.

Characterizing the genetic basis of X-linked hypertrichosis

In Chapter II.2 of my thesis, I investigated the genetic basis of the X-linked form of hypertrichosis in a family in which the Christiano lab previously reported linkage to chromosome xq24-27 over a decade ago (TADIN-STRAPPS *et al.* 2003). The initial approach searching for intragenic mutations using conventional sequencing strategies did not provide insight into the underlying genetic defect. With the advent of new technologies, we used a comprehensive approach combining high-resolution CNV analysis, FISH, and next-generation whole-genome sequencing technologies to detect the rare, pathogenic variant residing within the candidate linkage interval for this disease. Remarkably, this strategy enabled us to detect a large interchromosomal insertion with a complex sequence origin and inverted orientation that otherwise failed to detect using the conventional methods. We defined the precise location of this insertion at an extragenic palindromic site using whole-genome sequencing, and then

demonstrated complete cosegregation with the disease phenotype using a PCR-based screening approach, providing strong causal evidence for this insertion in XLH pathology (DESTEFANO *et al.* 2013). Importantly, our collaborations and access to several newer and more sophisticated methods (i.e. SOMA and whole-genome sequencing) empowered us with the tools to identify the critical insight in understanding the genetic architecture underlying XLH.

Three interchromosomal insertions occured at the same palindrome site

Considering the extremely rare occurrence of this condition in the general population, as well as the complex nature of the interchromosomal insertion resulting from eight double-strand breaks, the probability of identifying an independent, unrelated case characterized by the same unusual genetic mechanism is extremely low. Curiously, there are two additional XLH families of different ethnic backgrounds with linkage to the same genomic region (chrXq24-27), and recently, interchromosomal insertions were identified in both families (ZHU *et al.* 2011). Remarkably, all three insertions occurred at the exact same palindromic sequence on chrXq27.1. In all cases, at least one breakpoint resides within the palindrome site, and the insertion sequences are composed of autosomal CNV gains not present in control individuals or any public database. However, the sequence content and sizes of the insertions are distinct between the three families, verifying that they arose independently. Based on the nature of the genetic lesions identified in XLH, it was suggested that XLH is a genomic disorder, or one that is characterized by structural and numerical chromosome anomalies, though the precise mechanisms underlying the pathology remain unclear.

The identification of three distinct interchromosomal insertions at the exact same palindrome site raises intriguing questions, such as how do these interchromosomal insertions exert their pathogenic effects and cause excessive hair overgrowth? And what is the significance of their identical locations at the same palindromic site? To address the first question, I discuss several plausible mechanisms by which the insertions can act, including: 1. The disruption of a gene in the region; 2. The introduction of novel regulatory elements; 3. The presence of dosage-sensitive genes within the insertion; 4. The physical separation of neighboring genes from their regulatory elements, altering spatiotemporal expression; and 5. Disrupted chromosomal architecture and/or perturbed intra- and interchromosomal interactions required for normal development.

In all three families, the insertions are located at an extragenic palindromic site situated within a 485 Kb gene desert. In two families, both breakpoints reside within the center of the palindrome, and in the third family, the distal breakpoint resides within the palindrome, where the proximal is located 1.3 Kb from the palindrome (ZHU *et al.* 2011). Since all insertion events are contained within the 485 Kb gene desert, this provides evidence against the intragenic disruption hypothesis. Regarding the insertion content, the sequences are different among the three families (125 Kb from chr5q35.3, 300 Kb from chr4q31.2, and 389 Kb from chr6p21.2 (our study)), where each segment possesses different regulatory characteristics (based on ENCODE data). Furthermore, the genomic segments in two of the insertions (chr4 and 6) do not contain full genes, but rather fragments and/or introns, and none of these sequences have known roles in skin or hair follicle development. These observations highlight the unlikelihood of novel

regulatory elements and/or dosage-sensitive genes causing the XLH phenotype, and suggest that the presence of the insertion (rather than its content) mediates pathogenic effects.

In my thesis work described in Chapter II.2, I tested whether the interchromosomal insertion disrupted the expression levels of the genes in the surrounding region. In the following sections, I address the significance of the three insertions at the same palindromic sequence and overall incidence of interchromosomal insertions in sporadic conditions. Moreover, I discuss the rationale behind the approaches I used, the current questions that remain unanswered, as well as ongoing and future studies to elucidate the mechanisms by which the insertions act in XLH.

Palindromic sequences are identical in the 5'-3' orientation on one strand as they are on the other strand, thus, they are capable of forming hairpin structures that contain stem-loops (SMITH 2008). Interestingly, the cell recognizes these stem-loop structures in the same way (i.e. using the same machinery) as it does a Holliday junction (a branched structure of four double-stranded DNA molecules), and initiates a double-strand break. Studies in yeast have demonstrated that palindromic sequences as short as 140 bp can undergo conformational changes to form cruciform structures (substrates for the Holliday-junction resolvases to initiate double-strand breaks) (NAG AND KURST 1997). Thus, palindromic sequences are inherently unstable, and act as hotspots for meiotic recombination events in the genome (SMITH 2008). In the three XLH families, the insertion events occurred at a 180 bp palindromic site on chrXq27.1. Notably, deletions of varying sizes (both within and encompassing the palindrome) were found in several healthy individuals from PCR analysis and whole-genome sequencing studies,

highlighting the unstable nature of this sequence and its propensity for genomic rearrangement, insertions and deletions (ZHU *et al.* 2011).

The 180 bp palindrome on Xq27.1 is flanked on one side by a long-terminal repeat (LTR), as well as a long interspersed element (LINE). Comparative genomic studies of the orthologous genomic region in several non-human primates demonstrated that chimpanzee genome contains both halves of the palindrome, but they are in direct orientation and separated by the LTR, whereas the orangutan, rhesus and marmoset genomes only possess one half of the palindrome (ZHU *et al.* 2011). Therefore, the palindrome is considered evolutionarily recent since it is only found in humans. Importantly, while deletions involving the palindrome were detected in control individuals, interchromosomal insertions at this site were neither reported in healthy nor affected individuals with sporadic conditions, thus highlighting the rare and deleterious nature of the three insertions underlying XLH.

Interchromosomal insertions represent a distinct class of insertional translocations (ITs) that are rare events and require at least three independent chromosome breaks (KANG *et al.* 2010). Their estimated incidence in human pathologies was reported between 1:10,000-1:80,000 live births (FRYNS *et al.* 1984) (VAN HEMEL AND EUSSEN 2000). Subsequent studies using array CGH and confirmatory FISH in 18,000 clinical samples detected submicroscopic Its, increasing the overall incidence of ITs to 1:500, suggesting they may not be as rare as previously thought (KANG *et al.* 2010). Moreover, the identification of previously undetected ITs in reported CNV cases highlights the need for FISH to detect genomic location of CNVs identified. In Chapter II.2, we used this approach to identify the pathogenic variant in XLH.

Position effects

Based on the observation that the insertion content differs across the three XLH families, I postulated that the insertion may exert its pathogenic effects through the physical separation of neighboring genes from their regulatory elements (creating position effects), and/or disrupting chromosomal architecture and/or intra- and interchromosomal interactions required for normal development. In other forms of hypertrichosis our lab has studied, including Ambras syndrome and autosomal forms of CGHT, position effects on the *TRPS1* and *SOX9* genes were identified as a result of large chromosomal rearrangements, insertions and deletions as far as 7.3 Mb away (FANTAUZZO *et al.* 2008; FANTAUZZO *et al.* 2012). The Christiano lab and other groups have demonstrated that both of these genes play important roles in hair follicle morphogenesis, evidence by phenotypes in knockout mice (VIDAL *et al.* 2005; NOWAK *et al.* 2008; FANTAUZZO AND CHRISTIANO 2012; FANTAUZZO *et al.* 2012). Moreover, the position effect on *Trps1* is recapitulated in the *Koala* mouse model for hypertrichosis, where a chromosomal rearrangement upstream of *Trps1* led to decreased expression at the sites of pathology (FANTAUZZO *et al.* 2008), providing strong evidence for a position effect as the underlying genetic mechanism in Ambras syndrome (discussed in Chapter I.5).

In the context of XLH, it was suggested that the candidate gene affected by the interchromosomal insertions could be *SOX3*, a neural gene expressed in otic pladoes that lies 82 Kb telomeric to the insertion, given the known roles of *SOX2* and *SOX9* in regulating hair growth (VIDAL *et al.* 2005; NOWAK *et al.* 2008; CLAVEL *et al.* 2012). However, several human syndromes associated with *SOX3* mutations (i.e. X-linked

hypopituitarism, XX male sex reversal, X-linked recessive hypoparathyroidism, and premature ovary failure 1) do not manifest with hair phenotypes (BOWL *et al.* 2005; WOODS *et al.* 2005; FERREIRA *et al.* 2010; SUTTON *et al.* 2011). The expression of *SOX3* or any others in the surrounding region had not previously been tested in XLH, only in normal subjects, and *SOX3* is not expressed in the skin or hair follicle (ZHU *et al.* 2011). Interestingly, *SOX3* is expressed in the otic placodes that give rise to ears during development, and the XLH affected family members are deaf, suggesting that potential dysregulation of *SOX3* expression during morphogenesis may contribute to this pathological feature.

In my work, I examined the expression levels of the neighboring genes surrounding the 389 Kb insertion by qRT-PCR and RNA-seq on XLH patient and control samples. Where many of these genes, including SOX3 were not expressed at detectable levels in the skin or hair follicle, *FGF13* was significantly decreased in XLH patients. *FGF13* is an interesting candidate gene because of the known role of FGFs in regulating hair growth, specifically *FGF5,* where mutations cause the *angora* excessive hair overgrowth phenotype through an increased duration of anagen (HEBERT *et al.* 1994). Moreover, a previous report had demonstrated mouse *Fgf13* expression in the hair follicle bulge during all stages of the hair cycle (KAWANO *et al.* 2004), making this a plausible candidate gene for further investigation.

In both healthy and XLH hair follicles, I characterized the expression pattern of FGF13 using several methods of detection at both the mRNA and protein levels. Moreover, I analyzed Fgf13 localization in mouse tissues during morphogenesis and postnatal hair follicle growth and found that the same expression pattern is reflected in

human anagen hair follicles. We identified decreased levels of *FGF13* in human XLH and suggested this gene as one target of a position effect resulting from the interchromosomal insertion. This conclusion does not distinguish a direct from indirect interaction between the insertion and decreased *FGF13* levels, and does not eliminate the possible involvement of other gene(s) in the region. Recently, conditional deletion of *Fgf13* in mice demonstrated a role for Fgf13 as a microtubule-stabilizing protein required for neuronal polarization and migration (WU *et al.* 2012). However, global *Fgf13* loss-of-function and its consequence on the hair follicle remain unknown, since this gene was specifically ablated in neurons and hair growth was not examined in this model.

In humans, *FGF13* has been associated with X-linked mental retardation (XLMR) in patients with Borjeson-Forssman-Lehmann Syndrome, where it lies within the breakpoint of a duplication (GECZ *et al.* 1999). The behavioral abnormalities observed in *Fgf13* deficient mice recapitulate the developmental delay observed in humans with XLMR, suggesting that the *Fgf13* knockout mouse may be a suitable model for human XLMR. Interestingly, human patients with XLMR do not have reported hair or skin phenotypes, and the XLH patients reported in our study do not exhibit signs of XLMR. While it has been reported that position effects on genes can produce distinct phenotypes from the intragenic loss-of-function mutations, it is likely that reduced *FGF13* levels do not solely (if at all) contribute to the XLH pathology. In fact, our RNA-seq analysis revealed several genes required for normal hair growth (i.e. *FGF5*, *HOXC13*) whose expression levels were dysregulated in XLH, highlighting the complex downstream effects of the insertion on the expression of autosomal genes.

Considering the complex and rare nature of the interchromosomal insertion in XLH, the rarest of rare conditions, it is likely that the insertion mediates its pathogenic effects through a variety of mechanisms, including the disruption of important inter- and intrachromosomal interactions. This hypothesis has not yet been tested in XLH or any other forms of inherited hypertrichosis, since beyond functional studies in mouse models, no direct evidence for perturbed chromosomal interactions has been demonstrated. In my work, I performed an *in silico* analysis of regulatory elements at the Xq27.1 locus and observed several enhancer elements both telomeric and centromeric to the insertion, many of which were human-specific and several that demonstrated strong lacZ reporter activity in the developing mouse ectoderm (VISEL *et al.* 2007).

To clearly elucidate the chromosome interactome signature in the healthy hair follicle and in the context of hypertrichosis, a genome-wide chromosome conformation capture analysis using Hi-C (3C with high-throughput sequencing) could be performed in the future. This unbiased approach will reveal both inter- and intra-chromosomal interactions at the Xq27.1 locus that are normally active in hair follicle cells and potentially reveal novel interactions as a result of the interchromosomal insertion. Ideally, this experiment would be performed on cells from all three XLH families and in conjunction with transcriptome sequencing to more accurately define the global dysregulation of chromosome interactions in XLH. This analysis, combined with further functional studies employing cell culture assays and mouse models will shed mechanistic insight into the dysregulation of key regulators of hair follicle morphogenesis in XLH.

In summary, the goal of my work in Chapter II.2 was to uncover novel genetic mechanisms associated with a rare, sporadic condition affecting hair growth. Given the rare incidence of these syndromes within the general population, we used a comprehensive approach encompassing a wide range of methods to detect complex, sporadic lesions. Importantly, utilizing next-generation sequencing techniques provided us with innovative tools that proved essential to identify the pathogenic lesion. These and other newer technologies have profoundly improved and accelerated our ability to identify genetic mechanisms underlying human hair growth. Our downstream analyses uncovered a position effect on a distant gene, *FGF13*, whose expression levels were markedly reduced in patient hair follicles. Future functional studies at the genome-wide level will be instrumental in defining the precise mechanisms by which these interchromosomal insertions exert their pathological effects in XLH.

Chapter III

Characterizing single-gene effects on human hair growth

ABCA5 loss-of-function mutation in autosomal recessive hypertrichosis

(Manuscript #2)

Chapter III.1 Preface

The genetic architecture of rare, Mendelian disorders is generally composed of rare variants with large effects, and in familial cases where the underlying genetic lesions are enriched within a pedigree, the best approach to identify the candidate gene has been to perform linkage and haplotype analysis using genetic markers (i.e. microsatellites or SNPs). However, small families with one or few affected individuals are not well powered for linkage. Assuming the effect size is large (i.e. disrupting gene function), and the genetic lesion is identical by descent (IBD) in related affected or carrier individuals, the most effective strategy to identify the causal variants is to perform whole exome sequencing. In Mendelian disorders, studying consanguineous families has greatly facilitated gene discovery because rare, recessive variants with large effects are overrepresented in the gene pool relative to the general population, and thus, are readily detectable.

In recent years, whole-exome sequencing has emerged as the gold-standard for the study of rare, Mendelian disorders, both recessive and dominant, and has proven very useful for identifying novel mutations that cosegregate with the disease phenotype in familial cases. A candidate-based filtering approach is often used to identify the causal variant, followed by genotyping the related affected and unaffected individuals to test for cosegregation of the mutation. However, when there is no candidate gene available based alone on function or previous genetic association, identifying copy number variants (CNVs) that encompass the same genomic region in other patients and assessing gene expression in the relevant tissue for the pathology can be very useful tools in narrowing the list of potentially causal variants.

In Chapter III.2, I characterized single-gene effects on human hair growth by investigating the underlying genetic basis of autosomal recessive congenital generalized hypertrichosis terminalis (CGHT) (OMIM #135400). We ascertained a small consanguineous family from Yemen that was not sufficiently well powered for linkage analysis, however, given the consanguineous union between the parents of the proband, we assumed an autosomal recessive mode of inheritance. Drs. Kwame Anyane-Yeboa, Heather Feenstra, and Nanette Silverberg (St. Luke's-Roosevelt Hospital Center) performed clinical evaluations of the proband (discussed in Chapter II Materials and Methods) and obtained clinical photos as well as blood and punch biopsy samples for the subsequent molecular analysis and functional studies that I performed. Drs. Karl Gledhill and Zongyou Guo provided me with assistance in the initial isolation and culturing of primary keratinocytes and fibroblasts from patient skin biopsies.

Given the autosomal recessive mode of inheritance as a result of consanguinity, we postulated that underlying mutation resided within the exome. To identify rare, homozygous recessive variants with large effects, we performed whole-exome sequencing on the proband as well as on both parents through Ambry Genetics (Allso Viejo, CA) and applied a standard filtering approach using bioinformatics tools (discussed in Chapter III.2 Materials and Methods) to identify pathogenic variants. This revealed three candidate genes with homozygous mutations, all of which cosegregated with the disease phenotype but none were associated with a Mendelian disease. I next tested their gene expression levels in healthy human skin and hair follicle samples using RT-PCR, and analyzed their genomic positions to determine overlap with known CNVs. This approach successfully identified *ABCA5* as the candidate gene, as *ABCA5*

possesses the highest level of expression in human skin and hair follicles and most importantly, resides within the minimal common region (chr.17q23-24) to six reported CGHT cases (one of which is reported in Chapter III.2). Despite the identification of these CNVs, no intragenic mutations have previously been identified in CGHT cases. For these reasons, I selected *ABCA5* for further investigation, performing detailed molecular analyses and functional studies to characterize the novel homozygous recessive mutation in human CGHT.

Using PCR and Sanger sequencing, I performed cosegregation analysis and tested the overall consequence of the mutation on mRNA splicing. I next performed qRT-PCR and immunoblotting to determine mRNA and protein levels, respectively, in both keratinocytes and fibroblasts. To gain insight as to the specific cell type(s) synthesizing *ABCA5* in the human hair follicle, I performed immunohistochemistry, immunofluorescence staining, *in situ* hybridization, as well as RT-PCR on microdissected tissues of the skin and hair follicle (discussed in detail in Chapter III.2 Results and Materials and Methods). I next characterized the expression pattern of *Abca5* in mouse skin and hair follicles using immunohistochemistry and immunofluorescence analysis to determine whether it is conserved, and used the adult testis as a positive control for *Abca5* expression.

To evaluate the consequence of the *ABCA5* mutation in CGHT patient tissues, I performed immunoblotting on CGHT keratinocytes and fibroblasts as well as immunofluorescence staining on cryosectioned human hair follicles. Since ABCA5 is a reported glycosylated protein, I analyzed its glycosylated and unmodified forms in both healthy and affected tissues using PNGase, an enzyme that removes *N*-glycosyl

groups, followed by immunoblotting. I next sought to determine the functional consequence of *ABCA5* loss-of-function in CGHT at the subcellular level. Since previous studies using mouse models have demonstrated a role for Abca5 in lysosome function and cholesterol transport, I hypothesized that CGHT cells may possess lysosomal defects and/or perturbed distribution of free cholesterol. To directly test this hypothesis, I collaborated with Drs. Claudia Dall'Armi and Gil Di Paolo, who provided me with expertise, reagents and assistance with these experiments and analyses.

To address lysosome function in CGHT, I tested its role in autophagy by measuring autophagosome clearance. I performed immunofluorescence staining on control and CGHT patient keratinocytes for Lamp1 (lysosome marker), p62 (marker of autophagic cargo), and LC3 (an autophagosome marker) following treatment with Bafilmoycin A1, a proton pump inhibitor that blocks the acidification and clearance of autophagosomes. Dr. Dall'Armi performed confocal microscopy and all quantifications. To visualize free cholesterol in CGHT cells, we stained for filipin, a polyene macrolide antibiotic and antifungal that fluoresces in the presence of unesterified cholesterol, and Lamp1 to detect lysosomes. Dr. Dall'Armi performed confocal microscopy analysis on the patient and control keratinocytes.

In conjunction with these studies, we reported an independent, sporadic case of CGHT, where only the proband was affected. Using the approach and methods outlined in Chapter II to study rare, sporadic conditions, we sought to identify chromosomal defects (CNVs and/or structural variants). Drs. Mazen Kurban, Vaidehi Jobanputra, Dorothy Warburton, Brynn Levy, and Antonio Sobrino performed the cytogenetic analyses, including G-banding of chromosomes, FISH using BAC clones, sub-telomeric

FISH, and CNV analysis using whole-genome mapping (Affymetrix 500k and 2.7M arrays). Drs. Katherine Fantauzzo, Maija Kiuru, Marija Tadin-Strapps, and Anna Vitebsky were involved in the downstream cytogenetic analyses and functional studies, and patient materials and clinical photos from this case were obtained from Drs. Julio C. Salas-Alanis and Larissa Lopez-Cepeda. These cytogenetic analyses revealed a t3;17 translocation and cryptic 1.3 Mb deletion encompassing *ABCA5* on one allele. Given the similar phentoypes between sporadic and autosomal recessive CGHT, and the pathogenic association with *ABCA5*, I hypothesized that the second allele may harbor an intragenic mutation and performed Sanger sequencing to directly test this. I then sought to determine whether *ABCA5* levels were reduced in sporadic CGHT and used the same methods and approaches in my analysis of the autosomal recessive CGHT case to test *ABCA5* mRNA and protein levels.

Using a comprehensive and robust approach that spans multiple methods of detection has enabled the identification of a rare, novel splice site mutation that results in the loss-of-function of a gene not previously associated with a human disease. In Chapter III.3, I discuss the known functions of ABCA5, as well as future experiments to test the mechanism(s) by which it may regulate human hair growth.

OPEN ACCESS Freely available online

PLOS | GENETICS

Mutations in the Cholesterol Transporter Gene *ABCA5* Are Associated with Excessive Hair Overgrowth

Gina M. DeStefano[1], Mazen Kurban[2], Kwame Anyane-Yeboa[3], Claudia Dall'Armi[4,5], Gilbert Di Paolo[4,5], Heather Feenstra[6], Nanette Silverberg[6], Luis Rohena[3], Larissa D. López-Cepeda[7], Vaidehi Jobanputra[4], Katherine A. Fantauzzo[2], Maija Kiuru[2], Marija Tadin-Strapps[1], Antonio Sobrino[8], Anna Vitebsky[1], Dorothy Warburton[1,3], Brynn Levy[4], Julio C. Salas-Alanis[9], Angela M. Christiano[1,2]*

1 Department of Genetics and Development, Columbia University, New York, New York, United States of America, **2** Department of Dermatology, Columbia University, New York, New York, United States of America, **3** Department of Pediatrics, Columbia University Medical Center, New York, New York, United States of America, **4** Department of Pathology and Cell Biology, Columbia University, New York, New York, United States of America, **5** Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York, United States of America, **6** St. Luke's-Roosevelt Hospital Center, New York, New York, United States of America, **7** Centro Dermatológico Pascua, Mexico City, Mexico, **8** New York Presbyterian Hospital, New York, New York, United States of America, **9** Basic Science, Universidad de Monterrey, Nueva Leon, Mexico

Abstract

Inherited hypertrichoses are rare syndromes characterized by excessive hair growth that does not result from androgen stimulation, and are often associated with additional congenital abnormalities. In this study, we investigated the genetic defect in a case of autosomal recessive congenital generalized hypertrichosis terminalis (CGHT) (OMIM135400) using whole-exome sequencing. We identified a single base pair substitution in the 5′ donor splice site of intron 32 in the ABC lipid transporter gene *ABCA5* that leads to aberrant splicing of the transcript and a decrease in protein levels throughout patient hair follicles. The homozygous recessive disruption of *ABCA5* leads to reduced lysosome function, which results in an accumulation of autophagosomes, autophagosomal cargos as well as increased endolysosomal cholesterol in CGHT keratinocytes. In an unrelated sporadic case of CGHT, we identified a 1.3 Mb cryptic deletion of chr17q24.2-q24.3 encompassing *ABCA5* and found that *ABCA5* levels are dramatically reduced throughout patient hair follicles. Collectively, our findings support *ABCA5* as a gene underlying the CGHT phenotype and suggest a novel, previously unrecognized role for this gene in regulating hair growth.

Citation: DeStefano GM, Kurban M, Anyane-Yeboa K, Dall'Armi C, Di Paolo G, et al. (2014) Mutations in the Cholesterol Transporter Gene *ABCA5* Are Associated with Excessive Hair Overgrowth. PLoS Genet 10(5): e1004333. doi:10.1371/journal.pgen.1004333

Editor: Ruth Schmidt-Ullrich, Max-Delbrück-Center for Molecular Medicine, Germany

Received July 31, 2013; **Accepted** March 7, 2014; **Published** May 15, 2014

Funding: This work was supported by NIH/NIAMS Grant P30AR44535 (http://www.niams.nih.gov/) GMD was supported by Columbia University Department of Genetics and Development Training Grant (T32GM007088) and is now supported by an F31 fellowship (NIH/NIDCR; F31 DE023472-01A1). This work was supported in part by the Skin Disease Research Center in the Department of Dermatology at Columbia University (USPHS P3044345) and the NIH/NIAMS Grant R01AR44924 (to AMC). (http://www.niams.nih.gov/) The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: amc65@columbia.edu

Introduction

Inherited hypertrichosis, first described in the sixteenth century, is characterized by hair growth that is excessive for the body site and age of an individual and is independent of androgen stimulation [1], [2]. These syndromes are categorized as ectodermal dysplasias and are often associated with additional congenital defects, including cardiomyopathy, gingival hyperplasia, and craniofacial malformations [3]. The genetic basis of these syndromes remained largely elusive until recently, when our group and others reported several chromosomal rearrangements, copy number variants (CNVs) and position effects involving genes associated with autosomal dominant, recessive, sporadic, and X-linked forms of hypertrichosis [4–10].

We previously reported a position effect on the zinc finger transcription factor *TRPS1* in Ambras syndrome hypertrichosis in humans and the *Koala* hypertrichosis phenotype in mice, where *Trps1* expression was decreased at the sites of pathology for the phenotype [5]. More recently, we and others elucidated the genetic basis of X-linked hypertrichosis [9],[10], resulting from large interchromosomal insertions on the X chromosome. We found that a position effect occurs on a distant gene, *FGF13*, whose expression was markedly and selectively reduced in patient hair follicles, suggesting a novel role for this growth factor in hair follicle growth and cycling [10].

In the autosomal dominant form of CGHT, we identified a series of duplications of chromosome 17q24.2-q24.3 and reported a position effect on the *SOX9* gene, situated ~1 Mb downstream of these variants, and found *SOX9* expression was markedly reduced throughout the follicular epithelium of patient hair follicles [4]. In a separate report of CNVs in the same region of chromosome 17q24.2-q24.3, four CNVs were identified in several cases of CGHT [6], with an overlapping minimal region of 555 kb encompassing four genes: *ABCA6, ABCA10, ABCA5, and MAP2K6,* suggesting that disruption of one of these genes may contribute to the CGHT phenotype. Despite the identification of CNVs and/or position effects in this region of chromosome 17q24.2-q24.3, no

Author Summary

Inherited hypertrichoses represent a group of hair overgrowth syndromes that are extremely rare in humans and have remained an area of great interest to evolutionary geneticists since they are considered to be recurrences of an ancestral phenotype. These syndromes often present with additional congenital abnormalities including bone, heart and dental defects; thus, it is crucial to identify the mechanisms and genes underlying the pathology. Copy number variants (CNVs) have previously been reported in several cases of congenital generalized hypertrichosis terminalis (CGHT) with a minimal overlapping region of 555 kb encompassing four genes. However, no point mutations in these or any other single genes have been described to underlie the CGHT phenotype. In this study, we report the first loss-of-function mutation in an ABC transporter, *ABCA5* and identified an additional copy number variant in a separate case that lies within the minimal common region. We found high levels of *ABCA5* expression in both epithelial and mesenchymal compartments of human and mouse hair follicles, and in CGHT patients, this expression is significantly reduced or completely lost. ABCA5 is a lysosomal protein, and its loss-of-function compromises the integrity of lysosomes and leads to an intra-endolysosomal accumulation of cholesterol. Importantly, our findings support a novel role for *ABCA5* in regulating hair growth.

point mutations in these or any other single genes have been described to underlie the CGHT phenotype.

In this study, we investigated the genetic basis of autosomal recessive CGHT (OMIM135400) in a consanguineous family. Whole-exome sequencing revealed a novel, rare variant in the 5′ donor splice site of intron 32 of *ABCA5* that segregates with the phenotype in a homozygous recessive manner. *ABCA5* is highly expressed in human skin and hair follicles, and its expression pattern is conserved in mouse tissues as well. Importantly, the *ABCA5* loss-of-function mutation not only leads to a decrease in protein levels in both mesenchymal and epithelial compartments of CGHT patient hair follicles, but also to lysosomal dysfunction, which results in a defective clearance of autophagosomes under basal conditions and an overall accumulation of endolysosomal cholesterol in patient keratinocytes. We also identified an unrelated case of sporadic CGHT with a t3;17 translocation and cryptic 1.3 Mb deletion spanning *ABCA5*, and found that *ABCA5* levels were dramatically reduced in patient cells as well as throughout the hair follicle epithelium. Our findings implicate *ABCA5* as a gene with an essential role in hair growth.

Results

Clinical features, histology, and quantification of hair follicle length in congenital generalized hypertrichosis terminalis (CGHT)

We ascertained a proband from Yemen with CGHT segregating with gingival hyperplasia as well as epilepsy. Excessive hair growth was observed on the face, including the forehead, cheeks, and upper cutaneous lip, arms, upper and lower back and legs (Figure 1A). No other members of the patient's family are affected, however, the parents of the proband were consanguineous, suggesting an autosomal recessive mode of inheritance (Figure 1B).

Analysis of patient hair follicles obtained from a whole skin biopsy of the forearm by hematoxylin and eosin staining demonstrated that the hairs are of the terminal type since they were medullated, pigmented and penetrated deep into the dermis (Figure S1). Patient hair follicles were thicker than those of controls, and anagen hairs were present in the skin biopsy, whereas no anagen hairs were detected in control skin biopsies. We measured the length of the hair shafts from plucked patient and control forearm hair follicles and found that the patient hair follicles were significantly longer (87%), with an average length of 29.6 mm (±0.9 mm) compared to 15.9 mm (±0.6 mm) for control follicles ($P<0.0001$) (Figure 1C).

Whole-exome sequencing in CGHT identified a splice site mutation in *ABCA5*

To investigate the genetic basis of CGHT in this family, whole-exome sequencing was performed (Ambry Genetics) on genomic DNA obtained from the patient as well as both parents. Following sequencing, bioinformatics analysis (see Materials and Methods) and filtering on mode of inheritance ultimately lead to the identification of variants in 26 candidate genes (33 alterations) that were homozygous in the proband and heterozygous in both parents (Figure 1D). Further interpretative filtering based on literature searches focused on genotype-phenotype correlation revealed three candidate genes with homozygous mutations, *ABCA5, DGKZ,* and *ZNF253,* all of which are not currently associated with a Mendelian disease, and thus are considered novel. The nature of the homozygous mutations identified in these genes includes one splice site mutation (in *ABCA5:* c.4320+1G>C), one missense mutation (in *DGKZ:* c.1678C>T(p.P560S)), and one deletion (in *ZNF253:* c.429delA). Segregation analysis of these mutations revealed that both parents are heterozygous for all three mutations, whereas neither unaffected sister carries the *ABCA5* c.4320+1G>C and *DGKZ* c.1678C>T(p.P560S) mutations, and only one unaffected sister carries the *ZNF253* c.429delA mutation.

To determine whether these candidate genes are expressed in the skin and hair follicle, RT-PCR analysis was performed on RNA isolated from whole skin, which revealed that the *ABCA5* and *DGKZ* genes are abundantly expressed, suggesting a potential role for these two genes in the pathogenesis of the CGHT phenotype, whereas *ZNF253* is expressed at lower levels. The function of the *ZNF253* gene is unknown. The *DGKZ* gene encodes diacylglycerol (DAG) kinase zeta, a member of the DAG kinase family which phosphorylates DAG to phosphatidic acid and plays important roles in lipid signaling implicated in neurological diseases, including epilepsy, depression and Alzheimer's disease [11] – [13]. Moreover, mice deficient in the gene encoding DAG kinase, *epsilon (Dgke)* a member of the same gene family, exhibit features associated with epilepsy [14], suggesting the *DGKZ* substitution mutation may contribute to the pathogenesis of the patient's seizures. ABCA5, an ATP-binding cassette (ABC) protein, is a lipid transporter involved in the efflux of cellular cholesterol levels, and Abca5-deficient mice develop symptoms similar to several lysosomal diseases of the heart [15], [16]. Interestingly, *ABCA5* lies in the minimal common region to four reported familial cases and one sporadic case of autosomal dominant CGHT, both with and without gingival hyperplasia [4] – [6], suggesting that mutations in this gene may be associated with CGHT.

The *ABCA5* mutation results from a G-to-C substitution in the first nucleotide of intron 32 (Figure 1E). Sanger sequencing was performed on genomic DNA from the proband as well as the unaffected father, which confirmed homozygosity in the proband (II-2) and heterozygosity in the father (I-2) for the c.4320+1G>C mutation (Figure 1 E, F). Importantly, this mutation was not present in control individuals, determined by searching various genome databases and sequencing the genomic DNA of 10 control individuals. Moreover, a query for genetic variants that lie within

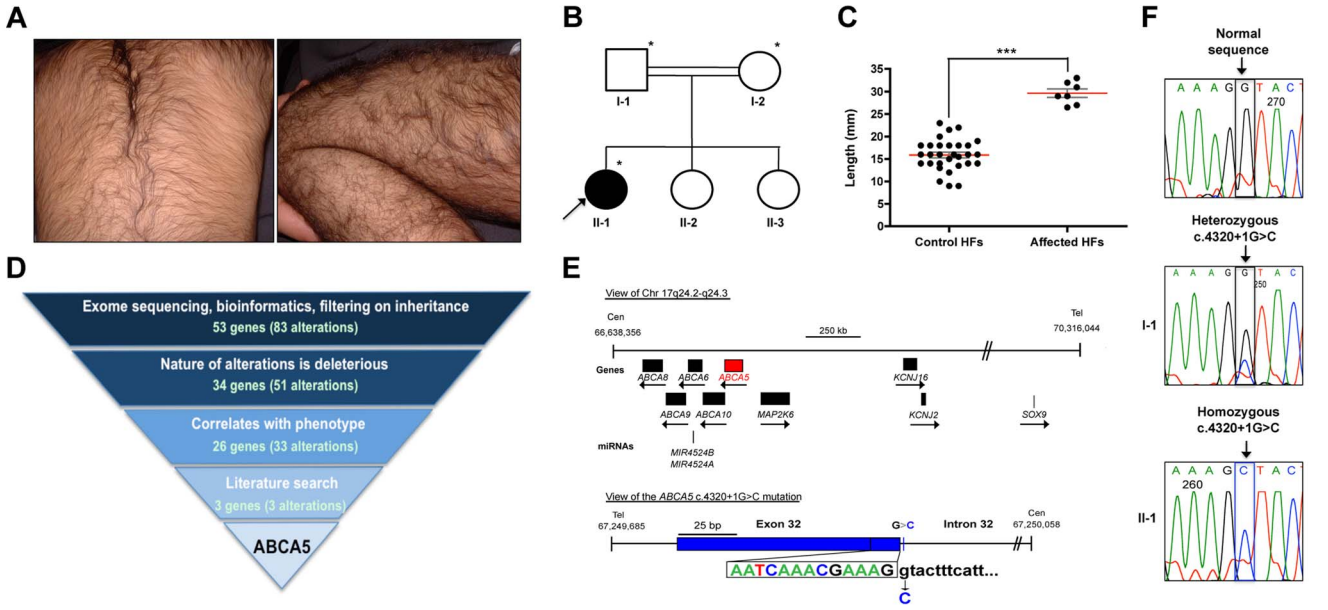


Figure 1. Whole-exome sequencing in a case of congenital generalized hypertrichosis terminalis (CGHT) revealed a splice site mutation in *ABCA5*. (A) Clinical photos of proband that illustrate the excessive hair overgrowth on the back (upper photo) and leg (lower photo). The proband is two years of age in this photo. (B) Pedigree of the family from a consanguineous union (indicated by double lines), where the proband (arrow) is the only affected member. Asterisks indicate the individuals on whom whole-exome sequencing was performed. (C) Quantification of the length (mm) of patient versus control hair follicles derived from the forearm revealed that patient hair follicles are 87% longer, with an average length of 29.6 mm (±0.9 mm) compared to 15.9 mm (±0.6 mm) for control follicles (***$P<0.0001$). (D) Whole-exome sequencing filtering strategy, including the number of alterations and genes obtained at each level. One alteration (c.4320+1G>C) was identified in the *ABCA5* gene, and cosegregation analysis in the family members revealed that both parents are carriers of this mutation, whereas the two unaffected sisters do not carry this mutation. (E) View of *ABCA5* (red) and the surrounding genes (black boxes) on chr17q24.2-q24.3, where arrows indicate direction of transcription. A magnified view of the c.4320+1G>C mutation in exon 32 of *ABCA5* with the exon 32-intron 32 boundary reference sequence annotated in colors corresponding to sequencing peaks. The G>C substitution is indicated below the control sequence. (F) Sequencing of control, carrier, and affected genomic DNA confirmed heterozygosity for the c.4320+1G>C mutation (boxed) in the father (I-1) and homozygosity in the proband (II-1). All coordinates reference the UCSC Genome Browser human reference genome hg19.
doi:10.1371/journal.pgen.1004333.g001

the *ABCA5* locus using the UCSC Human Genome (hg19) and Ensemble Genome Browsers verified that this variant is not a SNP or common variant associated with any human genetic disease. Thus, the c.4320+1G>C mutation in *ABCA5* is novel.

In light of the cosegregation of the mutation with the disease phenotype in the family, association with the CGHT phenotype in previously reported cases, and reported expression in the human hair follicle with *ABCA5* mRNA levels being the highest out of the four genes within the minimal common region [6], we further investigated the consequence of this mutation on *ABCA5* mRNA splicing and the potential role for this protein in hair follicle growth.

Sequencing of the mutant *ABCA5* transcript revealed aberrant splicing

In silico analysis of splicing events using the computational algorithms Berkley Drosophila Genome Project (BDGP) [17] and ESEfinder [18], [19] predicted that the c.4320+1G>C mutation results in complete loss of the donor splice site. To determine the consequence of the mutation at the transcript level, we amplified by RT-PCR the ~200 bp region flanking the mutation from patient, carrier, and control mRNA. Sanger sequencing revealed that the mutation leads to aberrant skipping of exon 32 in the proband (Figure 2A). As a result of joining exon 31 to exon 33, the remainder of the transcript downstream of the mutation is out-of-frame, leading to a premature termination 14 bp downstream of the mutation (Figure 2A). RT-PCR analysis of the exon 31–33 amplicon using RNA isolated from whole skin revealed the complete absence of the wild-type transcript in the proband. In contrast, we found the presence of the wild-type transcript at high levels and the mutant transcript at very low levels in the father (Figure 2A). The mutant mRNA most likely undergoes nonsense-mediated decay, since the mutation resides near the 3′ end of the transcript and the aberrant splicing event is predicted to affect the overall stability of the mRNA.

To investigate this possibility, we compared the relative *ABCA5* mRNA levels between the proband and unaffected carrier father by qRT-PCR using primers located at the 5′ end of the mRNA, and found that transcript levels were significantly reduced in patient whole skin (2.7-fold; $p<0.0001$), cultured keratinocytes (2.8-fold; $p=0.0016$), as well as fibroblasts (4.9-fold; $p<0.0001$), demonstrating that the mutant mRNA is unstable and undergoes nonsense-mediated decay (Figure 2B).

ABCA5 is abundantly expressed in the epithelial and mesenchymal lineages of human and mouse hair follicles

While *ABCA5* expression was previously identified in plucked human hair follicles by RT-PCR analysis [6], the specific cell type(s) synthesizing *ABCA5* in the hair follicle and potential expression in the surrounding dermal tissue remained unknown. Using *in situ* hybridization on human hair follicles in the growth phase of the hair cycle (anagen), we observed *ABCA5* transcript expression in both the epithelial and mesenchymal compartments, present within the outer root sheath (ORS) of the hair follicle as

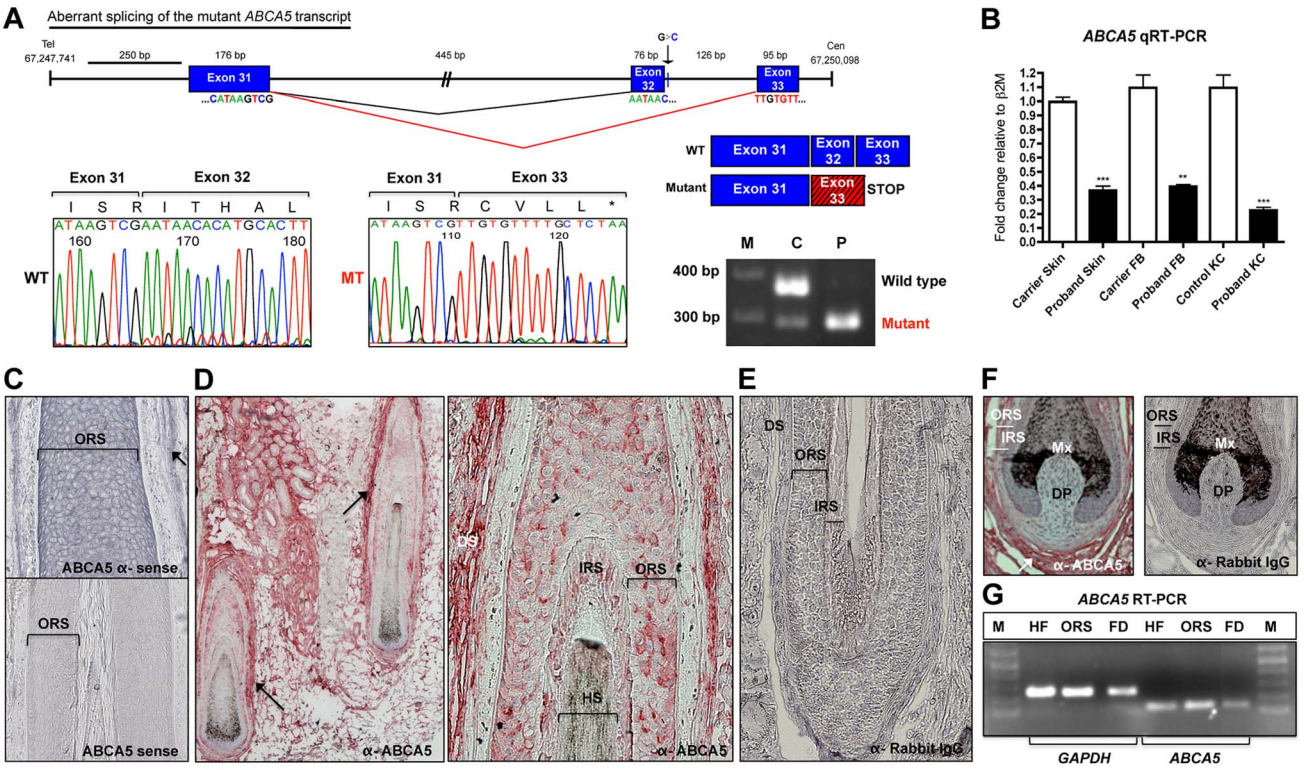


Figure 2. The *ABCA5* c.4320+1G>C mutation leads to aberrant splicing and nonsense mediated decay, and *ABCA5* is abundantly expressed in the skin and hair follicle. (A) Depiction of normal splicing between exons 31 and 32 (black lines) and aberrant skipping of exon 32 as a result of the c.4320+1G>C mutation (red lines), where the adjoining of exon 31 to 33 leads to premature termination during translation. The out-of-frame exon 33 sequence (red box) is indicated by the hatched lines, followed by a stop codon. Sequencing of the mutant transcript (MT) confirmed the aberrant splicing event, where the amino acid sequence is listed above and the stop codon is indicated by an asterisk. RT-PCR on RNA from whole skin followed by gel electrophoresis of the exon 31–33 amplicon demonstrates complete loss of the wild-type (WT) transcript in the patient (P), and very low levels of the mutant (MT) transcript in the carrier (C). (B) qRT-PCR of *ABCA5* transcript levels in the carrier and patient revealed a significant reduction of levels in patient whole skin, keratinocytes and fibroblasts. A Student unpaired t test was performed with a cutoff *P* value of 0.05 for statistical significance. (C) *ABCA5* is expressed at strong levels throughout the outer root sheath (ORS) and dermal sheath (arrow) of the human hair follicle by *in situ* hybridization using an antisense probe, whereas the sense probe produced minimal background signal. (D) ABCA5 localizes to the perifollicular dermis, dermal sheath, as well as the hair follicle ORS and inner root sheath (IRS) by immunohistochemistry on paraffin-embedded sections. (E) The anti-rabbit IgG primary antibody control produced no signal. (F) Immunohistochemistry of ABCA5 on hair follicle end bulbs revealed strong expression in the dermal sheath and perifolliclar dermis but no expression was observed in the dermal papilla. (G) *ABCA5* is endogenously expressed in plucked scalp hair follicles, microdissected ORS, as well as perifollicular dermis (including dermal sheath), determined by RT-PCR. Expression was normalized to *GAPDH* levels. M = marker.
doi:10.1371/journal.pgen.1004333.g002

well as dermal sheath (Figure 2C). To determine the localization pattern of ABCA5 protein in human skin and hair follicles, immunohistochemistry (IHC) was performed on paraffin-embedded skin sections and expression was most evident in the dermal sheath, perifollicular dermis, ORS, and IRS of hair follicles (Figure 2 D–F). Since the ABCA5 antibody is polyclonal, we validated endogenous *ABCA5* expression in the hair follicle as well as in the surrounding perifollicular dermis using RT-PCR on these tissues and observed strong *ABCA5* expression in plucked hair follicles (HF), microdissected ORS, as well as perifollicular dermis (FD) including the dermal sheath (Figure 2G).

To assess whether the human ABCA5 expression pattern was conserved in mouse skin and hair follicles, we first determined whether we could detect Abca5 immunoreactivity in a site of known *Abca5* expression using the same polyclonal antibody. In mice and rats, *Abca5* mRNA levels are abundant in the testis by Northern blotting and *in situ* hybridization [20], [21]. Using immunohistochemistry and immunofluorescence staining on adult mouse testis sections, we observed strong localization of Abca5 to the basal cells of the seminiferous tubules, interstitial cells consisting of Leydig cells (as previously reported in [21]), as well as the tunica albuginea (Figure S2 A–B, G). In the epididymis, we found very strong and specific localization of Abca5 to the connective tissue outlining the cylindrical epithelium in the corpus and cauda regions, including fibrocytes and smooth muscle cells, as well as within the basal and tall columnar cells of the corpus cylindrical epithelium (Figure S2 D–E, H–I). The control testis and epididymis sections (no primary antibody) yielded no signal (Figure S2 C, F). Importantly, we observed the same localization pattern of Abca5 in these tissues using two different fixatives; an organic solvent (methanol/acetone) and cross-linking agent (formalin), and our data are consistent with previous reports on *Abca5* mRNA expression [20], [21].

We next investigated the localization pattern of Abca5 in the mouse anagen hair follicle using immunofluorescence staining and immunohistochemistry, and observed high levels of Abca5 localization to the ORS and IRS of hair follicles (Figure S3). Expression was also observed in the dermal sheath and perifollicular dermis by the immunohistochemistry method (Figure S3 E–F), similar to what we observed in human hair follicles

(Figure 2D). Importantly, ABCA5 localization in the skin and hair follicle appears to be conserved between human and mouse, and its broad expression pattern spans multiple cell lineages, both within and surrounding the hair follicle. Collectively, these data suggest a prominent, evolutionarily conserved role for this transporter in regulating hair growth.

Homozygous loss-of-function of ABCA5, an *N*-glycosylated protein

To determine the consequence of the *ABCA5* c.4320+1G>C mutation at the protein level, we performed immunofluorescence staining on control and CGHT keratinocytes cultured from whole skin biopsies, and observed a striking reduction in ABCA5 localization (Figure 3A). Immunoblotting revealed the loss of a 215 kDa band corresponding to the full-length transporter in its glycosylated form, as well as a 187 kDa band, representing the unglycosylated form [15] (Figure 3B). While the full-length transporter was predominantly detected in keratinocytes, in fibroblasts, we observed the expression of a previously reported truncated variant [20] that produces a half-transporter and a ~100 kDa band. Importantly, the band representing this truncated variant was not detectable in CGHT patient fibroblasts (Figure S4). Since ABCA5 is a reported glycoprotein possessing multiple sites for *N*-glycosylation, we investigated whether the c.4320+1G>C mutation abolished the glycosylated form of the protein in patient fibroblasts by incubating total protein in the presence or absence of the PNGaseF enzyme that removes all *N*-glycosyl modifications. Following immunoblotting, we observed that the band corresponding to the ~100 kDa isoform that is absent in the proband represents the glycosylated form of the protein (Figure S4). Since glycosylation is an important post-translational modification crucial to the proper folding, stability, subcellular localization and/or even function of many lysosomal proteins, including ABC transporters [22], this finding suggests a loss-of-function of both the full- and half-transporters encoded by *ABCA5* in CGHT.

Lastly, to evaluate the consequence of the c.4320+1G>C mutation on ABCA5 protein localization at the tissue level, we performed immunofluorescence staining on patient and control hair follicles and observed a striking reduction of ABCA5 protein throughout the outer and inner root sheaths of patient hair follicles in catagen and anagen (Figure 3C). Importantly, loss of ABCA5 expression at the site of pathology for the phenotype further demonstrates that the c.4320+1G>C mutation results in a complete loss-of-function allele.

ABCA5 loss-of-function in CGHT keratinocytes perturbs endolysosomal function

Lysosomes have been reported to play a role in hair cycling, where mice deficient for lysosomal proteases, cathepsin L (Ctsl) and lysosomal acid phosphatase 2 (Acp2), have delayed progression through the hair cycle resulting in periodic hair loss that is characterized by hyperproliferation of epithelial cells in the hair follicle and basal layer of the epidermis [23] – [26]. In mice, Abca5-deficient cells have aberrant processing of autolysosomes and autophagosomes [15], suggesting a role in lysosome integrity and/or autophagy, a catabolic process of intracellular digestion and recycling of organelles [27]. To gain insight into whether CGHT patient cells possessed intrinsic autolysosomal and/or autophagic defects, we visualized key markers of these organelles at the cytological level. LC3, a well-established marker for autophagosomes, is converted from its cytosolic form (LC3-I) to a lipidated form (LC3-II) which is able to re-localize and bind specifically to autophagosomal membranes [27]. A large pool of LC3-II is degraded upon lysosome-autophagosome fusion, when the internal content of autophagosomes is destroyed by lysosomal hydrolases. Immunofluorescence staining against endogenous LC3 revealed an increased number of LC3-positive particles (*i.e.*, autophagosome-like structures) in the affected keratinocytes compared with control, suggesting defects in the autophagy pathway (Figure 4A).

In order to discriminate between higher levels of basal autophagy and a defective clearance of autophagosomes, bafilomycin A1 (BAF), a proton pump inhibitor that blocks the acidification of lysosomes and thus the clearance of autophagosomes, was added to growing cells 2 hours prior to fixation. BAF treatment caused a 2-fold increase in the number of LC3-positive puncta representing autophagosomes and autophagolysosomes in control keratinocytes, compared to untreated keratinocytes (Figure 4A, Figure S5A). In contrast, BAF treatment failed to significantly increase the number of LC3-positive particles in the patient keratinocytes compared to untreated cells (Figure 4A, Figure S5A). Furthermore, immunofluorescence staining against p62, a polyubiquitin binding protein as well as an autophagic cargo, revealed a strong accumulation of puncta in patient keratinocytes, further suggesting autophagy defects (Figure 4A, Figure S5B). Overall, these results indicate that *ABCA5* loss-of-function in CGHT causes defects in the autophagy pathway, and more specifically, point to an impairment in the clearance of autophagosomes and autophagic flux under basal conditions. These results also suggest that the mechanism underlying the autophagy defects caused by the *ABCA5* mutation is a decrease in lysosomal function.

Because a previous report has shown a role for ABCA5 in the efflux of cholesterol in macrophages [16], we next tested whether patient cells exhibit defects in the metabolism and/or transport of free cholesterol. Additionally, because ABCA5 is localized to the lysosomal compartment, its mutation may affect the handling of lipoprotein-derived cholesterol in the endolysosomal system, perhaps contributing to the dysfunction of these organelles in patient cells. In order to visualize cholesterol, cells were stained with filipin, a polyene macrolide antibiotic and antifungal that fluoresces and detects unesterified free cholesterol [28]. Remarkably, the mutation in *ABCA5* produced an increase in intracellular filipin staining compared to controls (Figure 4B, top panel). More specifically, filipin-positive puncta were observed mostly inside Lamp1-positive structures (Figure 4B, lower panel) suggesting an accumulation of free cholesterol in the lumen of endolysosomal organelles, likely in their intraluminal vesicles. These data suggest that ABCA5 controls the fate of lipoprotein-derived cholesterol and that its mutation alters the intracellular traffic of free cholesterol, somewhat reminiscent of phenotypes observed in lysosome storage disorders, such as Niemann Pick disease Type C [29].

Cytogenetic analyses and breakpoint mapping in a sporadic CGHT case revealed a t3;17 translocation and cryptic 1.3 Mb deletion encompassing *ABCA5*

At around the same time we identified the *ABCA5* c.4320+1G>C mutation in homozygous recessive CGHT, we independently studied this candidate region of chromosome 17 in a sporadic case of hypertrichosis. We ascertained a patient from Mexico with hypertrichosis universalis congenita (OMIM145700), whose parents (nonconsanguineous union) and siblings were unaffected. The patient exhibited excessive overgrowth of terminal hairs on the extremities, back, chest, and the face. Moreover, histological analysis of patient hair follicles obtained from a skin biopsy of the

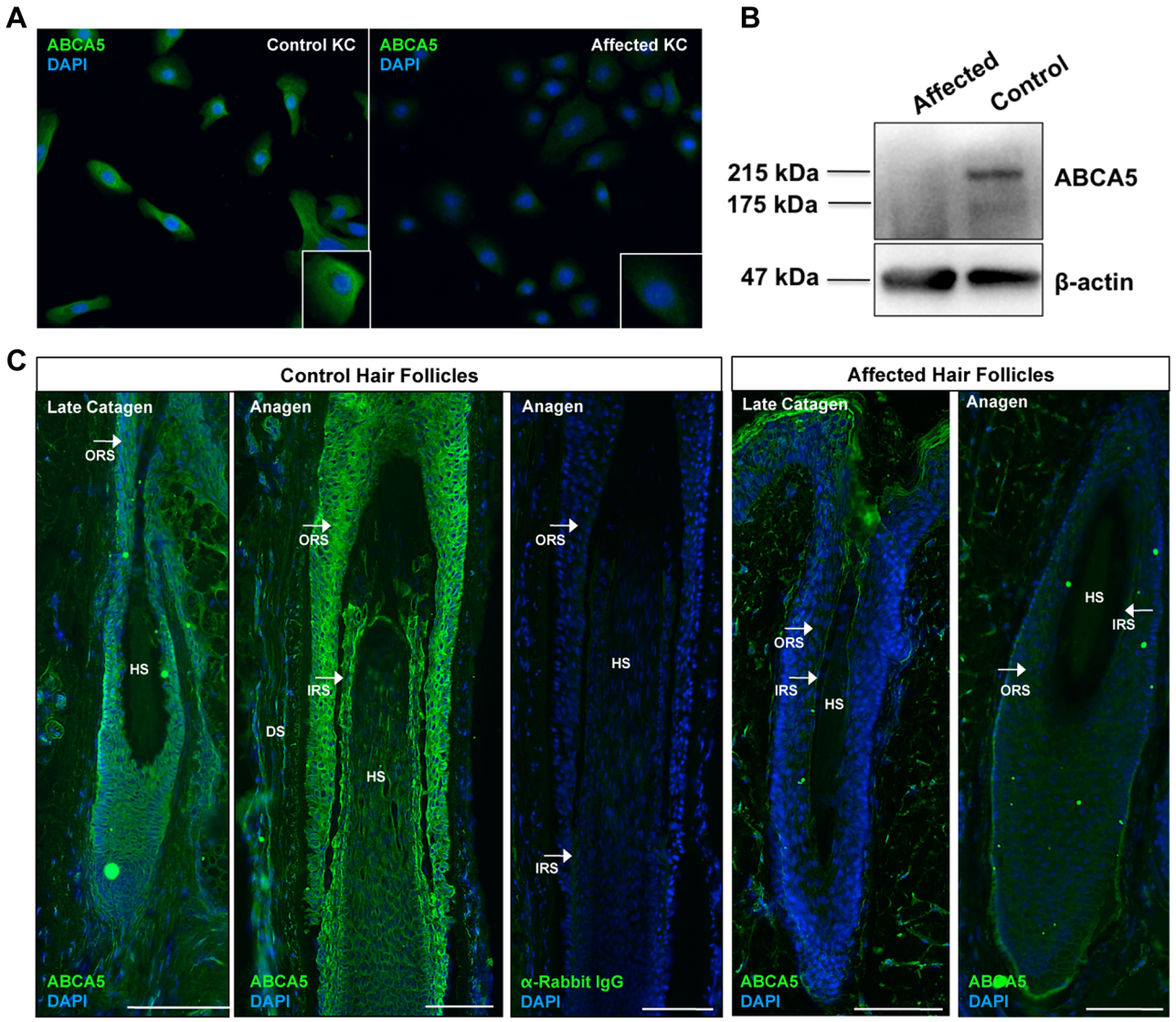


Figure 3. ABCA5 protein levels are significantly reduced in CGHT patient keratinocytes and hair follicles. (A) Immunofluorescence staining on cultured keratinocytes revealed a dramatic decrease in ABCA5 localization in CGHT keratinocytes compared to control (B) Immunoblotting on protein extracted from control and patient keratinocytes demonstrated loss of a 215 kDa band, which corresponds to the glycosylated form of the full-length ABCA5 protein, as well as a 187 kDa band, which is the unmodified protein in the patient compared to control. (C) Loss of ABCA5 localization to the outer root sheath (ORS) and inner root sheath (IRS) within patient hair follicles by immunofluorescence staining. Affected hair follicles in anagen and catagen were obtained from forearm skin biopsies, and control catagen hairs were obtained from forearm skin and anagen hairs from the occipital scalp. The anti-rabbit IgG primary antibody control produced no signal. HS = hair shaft; DP = dermal papilla; DS = dermal sheath.
doi:10.1371/journal.pgen.1004333.g003

lower back revealed the presence of large anagen hair follicles that penetrate deep within the dermis (Figure S6).

We initially performed karyotype analysis using G-banding of metaphase chromosomes that revealed a translocation between chromosomes 3q12 and 17q25, while the other chromosomes appeared cytogenetically normal. Chromosomal paint with two FISH probes against chromosomes 3 and 17 verified that no other chromosomes were involved in this rearrangement (Figure 5A), and telomere FISH (Figure S7A) confirmed that no subtelomeric sequences were lost as a result of the chromosome 17 breakpoint near the q telomere, suggesting an apparently balanced translocation event. To test whether this rearrangement segregated in the family, karyotype analysis was performed on both parents and siblings of the patient, but no abnormalities were detected, suggesting a *de novo* chromosomal rearrangement.

We next performed FISH analysis using BAC clones, which revealed a cryptic deletion at the breakpoint of chromosome 17, spanning BAC clones RP11-387O17 (chr17:66,354,485-66,568,819) and RP11-293K20 (chr17:67,076,187-67,261,438) (Figure S7B, C). To fine-map the deleted region, we utilized the Affymetrix 2.7M SNP array, which defined an 849.4 kb cryptic breakpoint deletion on chromosome 3p12.2 which did not contain any annotated genes, as well as a 1.3 Mb deletion on chromosome 17q24.2-q24.3 (Figure 5B) that encompassed 7 Ref Seq genes: members of the superfamily of ATP-binding cassette (ABC) transporters (*ABCA8, ABCA9, ABCA6, ABCA10, ABCA5*), mitogen-activated protein kinase kinase 6 (*MAP2K6*) and potassium

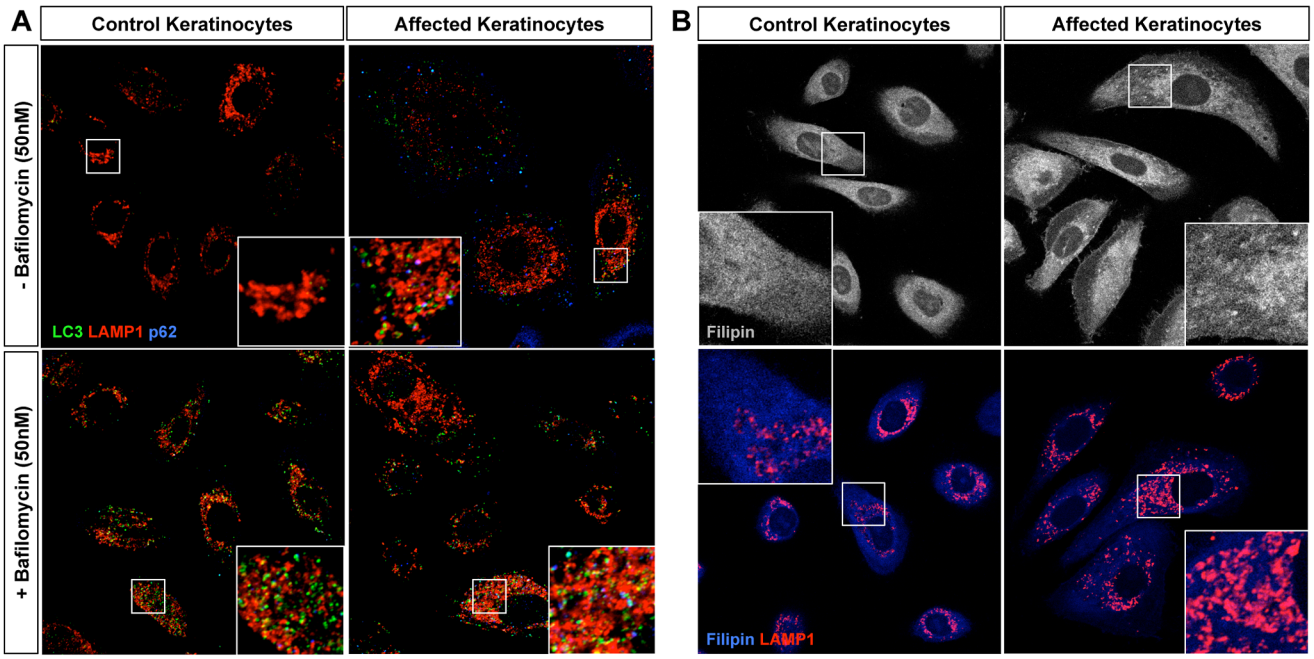


Figure 4. Homozygous loss-of-function of *ABCA5* perturbs lysosome function, resulting in an overall accumulation of autophagosomes and intracellular cholesterol levels in CGHT keratinocytes. (A) Control and CGHT keratinocytes were cultured in normal media in the presence or absence of 50 nM bafilomycin, fixed and immunostained. Confocal analysis of LC3 (green), Lamp1 (red) and p62 (blue) revealed an increased number of LC3-positive particles as well strong accumulation of p62 particles in affected keratinocytes. BAF-treated control cells possessed a two-fold increase in the number of LC3 puncta, whereas no significant difference was observed in BAF-treated affected cells (quantified in Figure S5). (B) Control and CGHT keratinocytes were fixed and immunostained for Filipin and Lamp1 (red), which revealed an accumulation and redistribution of free cholesterol to Lamp1-positive organelles in affected keratinocytes.
doi:10.1371/journal.pgen.1004333.g004

inwardly-rectifying channel J16 (*KCNJ16*). We confirmed the 1.3 Mb deletion with quantitative PCR (qPCR) on patient and control genomic DNA using primers specifically designed to amplify sequences flanking and within the deleted region (Figure 5C). To determine whether the translocation event resulted in the disruption of a gene, we searched for transcripts that mapped to the vicinity of the breakpoints using the Genome Browser available at the UCSC database (hg19) and subsequently cloned the breakpoints, which did not reveal any evidence for gene disruption.

The cryptic 1.3 Mb deletion at chr17q24.2-24.3 leads to a position effect on the downstream gene, *SOX9*

Several cases of CGHT have been reported in which CNVs on chr17q24.2-24.3 that lie within a 2.4 Mb region (Figure S8) have been identified and hypothesized to cause a position effect on the *SOX9* gene, a well-defined regulator of hair follicle stem cells [30], [31], which resides 1-2 Mb downstream of these variants. In our previous report of an autosomal dominant CGHT case, we found *SOX9* expression was dramatically reduced throughout the hair follicles of patients who possess duplications within the chr17q24.2-24.3 region [4]. To investigate the possibility that the 1.3 Mb deletion identified in the present study led to altered *SOX9* expression, we performed qRT-PCR on RNA isolated from patient and control keratinocytes and observed decreased *SOX9* expression in the patient keratinocytes by approximately 2.5-fold ($P>0.001$) (Figure 5D). This result is consistent with our previous findings, suggesting a contribution of decreased *SOX9* expression levels to the excessive hair overgrowth phenotype of CGHT. Given the association of SOX9 with CGHT and its known role in regulating hair follicle stem cells and outer root sheath differentiation [30], [31], we investigated the consequence of *ABCA5* loss-of-function on *SOX9* transcripts in the autosomal recessive CGHT case. Using qRT-PCR on patient and control keratinocytes, we did not detect a significant difference in *SOX9* gene expression (data not shown). While it remains possible that *SOX9* acts upstream of *ABCA5*, these genes may reside in separate pathways and utilize distinct mechanisms to contribute to the CGHT pathology.

ABCA5 mRNA and protein levels are dramatically reduced in both epithelial and mesenchymal compartments of the skin and hair follicle in sporadic CGHT

Since the 1.3 Mb deletion we identified encompasses the *ABCA5* locus, and both sporadic and autosomal recessive CGHT patients possess excessive overgrowth of terminal hairs, we next investigated the possibility that the second allele in the sporadic CGHT case may harbor a mutation in the *ABCA5* gene. Therefore, we sequenced all the exons of *ABCA5* and ~2 kb upstream of the gene, but did not find any mutations (data not shown). To test *ABCA5* mRNA expression levels in the sporadic CGHT case, we performed qRT-PCR in keratinocytes and fibroblasts cultured from a patient skin biopsy compared to control, and observed a 3.7-fold decrease in *ABCA5* transcript levels in patient keratinocytes ($p<0.0001$), and a 2.4-fold decrease in patient fibroblasts ($p<0.0001$) (Figure 6A). Immunoblotting in fibroblasts cultured from patient and control skin biopsies revealed a striking decrease in ABCA5 protein levels, and treatment with the *N*-glycosylase PNGaseF revealed loss of the glycosylated form of the protein (Figure 6B). Moreover, we found using immunofluorescence staining on patient and control hair follicles in catagen and anagen that ABCA5 protein levels were

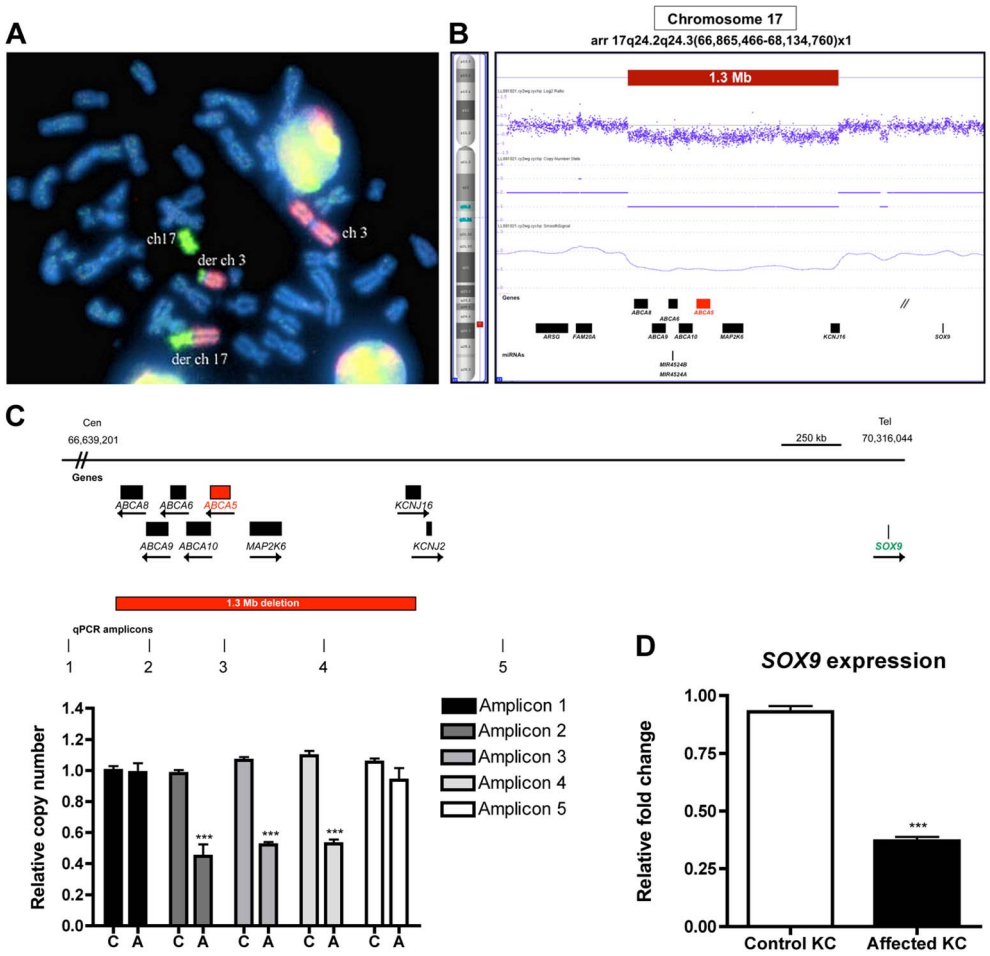


Figure 5. Cytogenetic analyses, breakpoint mapping, and copy number variant analysis in a sporadic case of CGH revealed a t3;17 translocation that leads to a cryptic 1.3 Mb deletion of chr17q24.2-24.3, and *SOX9* expression is reduced in patient keratinocytes. (A) Chromosomal paint was performed in a sporadic case of CGHT with commercially available whole chromosome probes for both chromosomes 3 and 17. Note: the probes do not bind to any of the other chromosomes except 3 and 17, and the derivative chromosomes suggesting that this is an isolated rearrangement between these two chromosomes. DAPI staining was used to visualize the entire chromosomal set (blue). (B) Copy number variant analysis using the Affymetrix Cytogenetics Whole Genome 2.7M array validated the cryptic deletion that spans 1.3 Mb (red box). RefSeq genes are depicted as black boxes. (C) Region of chr17q24.2-q24.3, including the genes and 1.3 Mb deletion identified by FISH and CNV analysis. The quantitative PCR (qPCR) amplicons are indicated as vertical lines, with three amplicons residing within and two flanking the deletion. qPCR was performed to confirm the deletion on control (C) and affected (A) genomic DNA using relative quantification and the primers listed in the Materials and Methods. (D) qRT-PCR of *SOX9* on RNA from control and affected keratinocytes revealed a 2.5-fold decrease ($P>0.001$) in expression in patient keratinocytes. A Student t test (unpaired) was performed with a cutoff *P* value of 0.05 for statistical significance. All coordinates reference the UCSC Genome Browser human reference genome hg19.
doi:10.1371/journal.pgen.1004333.g005

dramatically reduced throughout the outer and inner root sheath of the patient hair follicles (Figure 6C). Collectively, these results suggest that loss of one genomic copy of *ABCA5* and its surrounding regulatory elements severely disrupts expression from the other allele, and significantly reduces expression levels, suggesting haploinsufficiency of *ABCA5* in the sporadic case of CGHT.

Discussion

In this study, we investigated the genetic basis of a case of CGHT using whole-exome sequencing and identified a homozygous recessive loss-of-function mutation in a donor splice site of *ABCA5* that cosegregates with the phenotype. The c.4320+1G>C mutation leads to loss of *ABCA5* expression and localization within patient keratinocytes, fibroblasts and hair follicles compared to controls. We found that endogenous *ABCA5* is highly expressed in both epithelial and mesenchymal compartments of the hair follicle, and that its homozygous recessive loss-of-function is associated with defective lysosomes, as well as an overall accumulation of autophagosomes and endolysosomal cholesterol. In a case of sporadic CGHT, we performed detailed cytogenetic analyses, breakpoint mapping, CNV analysis and expression studies, which revealed a 1.3 Mb deletion encompassing the *ABCA5* locus and markedly reduced *ABCA5* levels in patient cells and hair follicles.

In inherited hypertrichoses, the mechanism(s) that lead to an excessive hair overgrowth phenotype observed include an increased anagen duration, increased hair follicle density, as well as a vellus-to-terminal transformation, where fine, unpigmented and unmedullated hairs are stimulated to grow deeper into the dermis, widen, and become medullated as well as pigmented [1], [2]. In the case of CGHT, a condition that is present at birth,

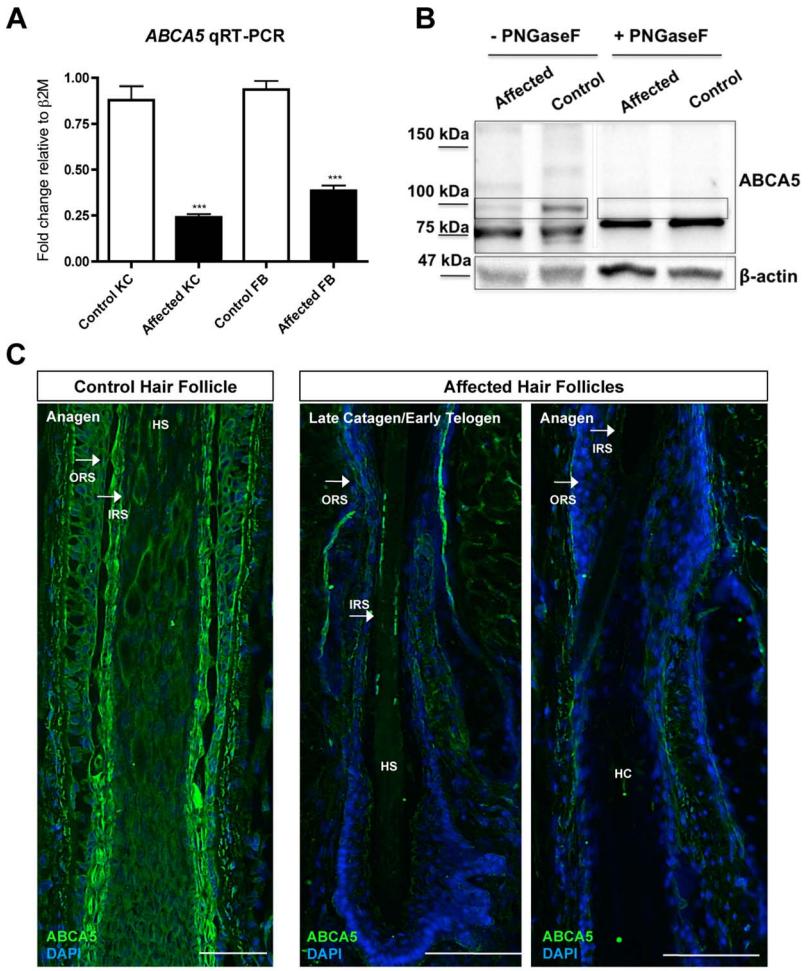


Figure 6. *ABCA5* levels are markedly reduced in patient hair follicles of the sporadic CGHT case. (A) qRT-PCR for *ABCA5* on control and affected keratinocytes and fibroblasts cultured from whole skin biopsies revealed a significant reduction in the patient cells compared to controls. A Student unpaired t test was performed with a cutoff *P* value of 0.05 for statistical significance. (B) Immunoblotting on protein extracted from carrier and patient fibroblasts in the presence or absence of the enzyme, PNGaseF revealed loss of a ~100 kDa band that is the glycosylated form of the protein in patient fibroblasts relative to controls. (C) Immunofluorescence staining on control and patient hair follicles in the catagen and anagen stages demonstrated reduced ABCA5 localization throughout the outer root sheath (ORS) and inner root sheath (IRS) of patient hair follicles (compare to control catagen hair follicle in Figure 3C). HS = hair shaft; HC = hair canal.
doi:10.1371/journal.pgen.1004333.g006

terminal hairs develop at sites of the body where vellus hairs should be, a developmental event that most likely occurs during late hair follicle morphogenesis or immediately following. Lanugo hairs shed *in utero* and then replaced by vellus hairs are instead replaced with terminal hairs. In addition to a vellus-to-terminal transformation, it is likely that the excessive hair overgrowth phenotype in CGHT is also contributed by an increased duration of anagen, as the hairs produced across the body are longer than what is considered normal for the age, race, and gender of the individual. In the autosomal recessive CGHT case in this study, we found that the length of hair follicles plucked from the forearm was significantly greater than that of control hair follicles. Moreover, histological analysis of patient skin biopsies from extensor skin of the extremities revealed the presence of large anagen hair follicles situated deep in the dermis, whereas hair follicles in anagen were not present in control skin biopsies.

Using several methods of detection, we found that ABCA5 is strongly expressed in both mesenchymal and epithelial lineages of the human hair follicle. In mouse hair follicles, we observed the same expression pattern as in human follicles, and using immunostaining with a polyclonal ABCA5 antibody, detected Abca5 localization in the testis and epididymis—positive sites of Abca5 mRNA expression. In the skin, the mesenchymal expression (perifollicular dermis, dermal sheath) appears more intense using formalin-fixed paraffin embedded (FFPE) samples, and the epithelial outer and inner root sheath expression is more intense using frozen sections. These differences are most likely artifacts of antigen accessibility and tissue preservation (organic solvent vs. cross-linking agent) between the two methods (i.e. dermal and subcutaneous tissue is better preserved in the FFPE sections). However, RT-PCR analysis and *in situ* hybridization demonstrated strong *ABCA5* expression in both epithelial and mesenchymal lineages, and negative controls for both mRNA and protein studies yielded no signal.

High levels of ABCA5 localization to the hair follicle IRS is intriguing, as several genes involving lipid metabolism or transport possess strong inner root sheath expression and have been described to underlie human genetic conditions with sparse hair. *P2RY5*, encoding an orphan G protein-coupled receptor, is mutated in patients with autosomal recessive wooly hair syndrome

[32], [33]. Likewise, *LIPH* encodes the phospholipase A1 family member that converts phosphatidic acid into lysophosphatidic acid, and is mutated in patients with autosomal recessive wooly hair and hypotrichosis [32] – [34]. While the precise mechanism by which these genes control hair growth remains elusive, we postulate that the loss of a functional ABCA5 transporter in the IRS of patient hair follicles may contribute to excessive hair overgrowth through blocking cholesterol efflux.

Human *ABCA5* is also highly expressed in perifollicular dermal tissue, including the dermal sheath of the hair follicle, and the number of dermal papilla cells has been shown to be important in determining hair follicle size in mice [35]. Since patient hair follicles are larger and longer, we postulate that they may contain an increased number of dermal papilla cells, but were unable to directly test this using the skin biopsies we received. Additionally, the control of hair follicle size is linked to BMP levels in the skin, where mice overexpressing the BMP antagonist, Noggin, have enlarged anagen hair follicles [36]. Therefore, ABCA5 transport of cholesterol may normally limit hair growth and intersect with the BMP signaling pathway, such that in *ABCA5*-deficient hair follicles, the excess buildup of intra-endolysosomal cholesterol attenuates BMP signaling and recapitulates the enlarged hair follicle phenotype of the Noggin overexpression mouse model. Functional studies using *Abca5*$^{-/-}$ mice would be required to directly address this.

ABCA5 has not previously been described to be associated with a human condition, particularly one manifesting a hair or skin phenotype. In humans, ABCA5 is a reported biomarker for tumor stem cells in osteosarcoma based on its overexpression, where it is also highly expressed in melanoma and undifferentiated colon and ovarian carcinomas [37] – [39]. Consistent with the reported role of ABC transporters in tumor biology conferring resistance to drugs and chemotherapy/cancer-related substrates *(i.e.* phospholipids and cholesterol) [40], *ABCA5* upregulation may protect undifferentiated tumor cells from an intracellular accumulation of cholesterol and other sterols.

Mutations in ABC transporters have been reported in several other genetic skin diseases, including *ABCA12* in lamellar ichthyosis type 2 as well as the severe and lethal ichthyosis, Harlequin ichthyosis, *ABCC9* in Cantú syndrome, and *ABCC6* in the connective tissue disorder pseudoxanthoma elasticum (PXE) [41] – [46]. *ABCA12* is expressed in the skin and its deficiency leads to abnormal lipid transport in keratinocytes, effectively compromising the skin barrier in a cell-autonomous manner [46]. Recently, dominant missense mutations have been reported in the ABC transporter gene, *ABCC9* in several cases of Cantú syndrome, a form of congenital hypertrichosis segregating with osteochondrodysplasia, distinctive facial features, as well as cardiac defects [44], [45]. The *ABCC9* gene encodes a transmembrane protein that is a constituent of a potassium channel complex, and electrophysiology experiments revealed that the missense mutations have gain-of-function or activating effects on the protein, as they reduce ATP-dependent inhibition of the channel [44]. Interestingly, the *ABCC6* gene encodes the multidrug resistance associated protein 6 (MRP6), and is exclusively expressed in the liver and kidney, suggesting that mutations lead to ectopic mineralization within the skin of PXE patients via a non cell-autonomous mechanism [47]. Importantly, this suggests that blocking the allocrite, or transported substrate, is the key determinant of the phenotype, because *ABCC6* itself is not expressed in the skin, yet its loss-of-function manifests with a cutaneous phenotype.

The allocrite of ABCA5 has remained unknown, however, we have observed a cholesterol transport defect in human CGHT keratinocytes and mouse Abca5 has previously been reported to play a role in cholesterol efflux in macrophages [16]. In rats and now mice, Abca5 localizes to the Leydig cells of the testis, a primary site for cholesterol processing and sterol hormone synthesis [21], suggesting cholesterol is a potential allocrite for this transporter. Importantly, these reports as well as the subcellular localization pattern of Abca5 are consistent with our observation that *ABCA5* deficiency in CGHT patient cells causes a redistribution of free cholesterol and accumulation of this sterol in the endolysosomal compartment. The resulting sequestration of cholesterol within these organelles may prevent the delivery of this essential lipid to other cellular compartments, such as the plasma membrane, altering their physical and signaling properties. Alternatively, accumulation of endolysosomal cholesterol may decrease the overall fitness of these organelles, leading to a decrease in the degradative capacity of lysosomes and causing secondary defects in autophagy.

Cholesterol as a covalent modification is required for the biological activity of several signaling proteins in the Wnt and Shh pathways that regulate hair follicle morphogenesis and cycling [48]. Overexpression and knockout mouse models for such molecules have demonstrated critical roles in hair follicle patterning and possess phenotypes reminiscent of hypertrichosis. For example, mice deficient for the *Smoothened (Smo)* gene in *Keratin-14*-expressing epithelia develop *de novo* and abnormally large hair follicles from increased mesenchymal SHH signaling [49], and the overexpression of a stabilized beta-catenin *(Ctnnb1)* induces early placode formation as well as an increased density of hair follicles [50]. We postulate that decreased free cholesterol in *ABCA5* deficient CGHT cells may modulate the range and/or morphogenetic activity of such proteins either in the epithelium or mesenchyme during hair follicle patterning. Future experiments employing lineage-specific deletions of ABCA transporters (i.e. *Abca1, Abca5*) will be useful in testing this, where free cholesterol levels can be measured biochemically and by filipin immunostaining, followed by a thorough analysis of hair follicle morphology and density. Similarly, a pharmacological approach to deplete sterols (i.e. cyclodextrin treatment) or block ABCA transporter activity (i.e. glyburide treatment) in pregnant dams during hair follicle morphogenesis may provide insight into the role of cholesterol transport in hair patterning and growth.

In mice, the biogenesis of cholesterol and other lipids has been shown to play a role in regulating hair growth, as mice deficient for *Stearyl-CoA Desaturase I (SCD1)*, the enzyme required for the biosynthesis of monounsaturated fatty acids that compose cholesterol esters as well as membrane phospholipids and triglycerides, develop alopecia and possess a sparse hair coat with dry, scaly skin [51] – [53]. Furthermore, topical treatment of mouse skin with a cholesterol biosynthesis inhibitor of a sterol precursor leads to hair loss through the induction of catagen [54]. It remains to be determined whether *Abca5*$^{-/-}$ mice possess a hair phenotype, since this has not been directly examined and hair phenotypes in mutant mice deficient for several factors can be subtle in the absence of detailed histological analysis [55]. If hair patterning is aberrant in *Abca5*$^{-/-}$ mice, lineage-specific deletions of *Abca5* followed by morphometric and densitometric analysis of hair follicles may provide insight into the target cell type underlying the pathology in CGHT.

In the sporadic CGHT case, we found that *ABCA5* transcripts were significantly reduced. One possible explanation for this is that a *trans* mechanism for regulating *ABCA5* gene expression via an enhancer or some other regulatory element is in place, such that deletion of one copy of *ABCA5* and its regulatory elements abrogates gene expression on the other allele. Conversely, *ABCA5*

may be imprinted or expressed preferentially from one allele, in which case, deleting the transcriptionally-active allele would result in loss of gene expression, thereby recapitulating the phenotype of the loss-of-function mutation. Alternatively, *ABCA5* may be a dosage-sensitive gene, where altering the copy number is sufficient to produce a phenotype, but perhaps in a distinct manner from the loss-of-function mutation identified in the present study, since carriers are unaffected. Since we observed a decrease in *SOX9* expression in sporadic CGHT patient keratinocytes, consistent with our previous findings in an autosomal dominant CGHT case [4] and considering the known role of SOX9 in hair follicle stem cells [30], [31], it is likely that the excessive hair overgrowth phenotype observed in sporadic CGHT may be the result of reduced levels of both *ABCA5* and *SOX9* transcripts. We did not detect altered *SOX9* expression in the autosomal recessive CGHT case, suggesting either that *SOX9* acts upstream of *ABCA5* or that the two genes reside in separate pathways, employing distinct mechanisms to contribute to the CGHT pathology.

Importantly, this is the first report of the *ABCA5* gene to be associated with a human genetic condition, and the first reported case of CGHT in which a point mutation in a single gene contributes to the excessive hair overgrowth phenotype. Collectively, our findings provide insight into the mechanisms by which ABCA5 may directly or indirectly act to control hair growth at the subcellular level. While there is evidence to support a role for *ABCA5* in both epithelial and mesenchymal compartments of the hair follicle, the precise cell type(s) deficient for *ABCA5* that is associated with excessive hair overgrowth remains unclear. In light of our findings, we postulate that the loss-of-function of *ABCA5* during morphogenesis abrogates lysosome function and cholesterol efflux in hair follicle cells, leading to the aberrant activity of one or more downstream signaling pathways with crucial, defined roles in hair follicle induction, development and cycling.

Materials and Methods

Ethics statement

Informed consent was obtained from all subjects and approval for this study was provided by the Institutional Review Board of Columbia University in accordance with the Declaration of Helsinki Principles.

Patient information and materials

In the autosomal recessive CGHT case, the proband is an eleven-year-old girl with congenital generalized hypertrichosis segregating with gingival hyperplasia and epilepsy. She was the product of a full-term uncomplicated pregnancy to a then 18-year old Yemeni mother whose husband is her first cousin. The patient's initial genetic evaluation was at 6 months of age; chromosome analysis revealed 46, XX normal complement and SNP arrays revealed multiple regions of homozygosity consistent with parental consanguinity. Severe hypertrichosis and a low anterior hairline that merged with the eyebrows were observed, and excessive body hair was present on the patient's back and extremities, as well as on the upper lip, genitalia, and axillary regions. Hair overgrowth progressed on her face trunk, eyebrows and eyelashes, and scalp as well as body hair is long and coarse. The gingival hyperplasia became evident at 18 months of age and progressed to the extent that it interfered with tooth eruption and feeding. Partial gum surgical resection was initially performed at 3 years of age and has been performed several additional times over the past eight years due to excessive gum overgrowth. Extensive endocrine and metabolic workups have been performed over the years and all results have been normal.

In the sporadic CGHT case, the patient initially presented to the clinic at the age of three months. On physical exam, she had universal hypertrichosis that was accentuated on the extremities, the back, chest, and the face. As expected, regions that are devoid of hair follicles including the palms, soles and distal phalanges were spared. In addition, the patient had unusual facial features, including a broad base of the nose, widely set eyes, and bulbous tip of the nose. The parents reported that the excessive body hair was shed during the first month of life, but then grew back with increased length. The mother did not report any drug intake, major illness or trauma during pregnancy, and the girl was born at full term. There is no history of hypertrichosis in the family. Extensive physical examination, radiological studies including X-rays (skull, long bones and hands), laboratory evaluations (complete blood count, blood chemistry, general urine examinations) and abdominal ultrasonography did not reveal any abnormalities.

In the autosomal recessive CGHT case, two small punch biopsies from extensor skin of the extremities, 3–5 cc of blood, in addition to plucked hairs from the forearm were received from the proband (II-1) and the unaffected father (I-1). For each individual, one biopsy was divided into two pieces: one flash frozen for RNA extraction and the other embedded in OCT for histological and immunological analyses. The other piece of whole skin was used for the culturing of keratinocytes and fibroblasts. In the sporadic CGHT case, a skin biopsy from the back of the affected individual was obtained; 3–5 cc of blood for DNA and RNA extraction was also obtained from the patient and her unaffected parents. Control hair follicles used in expression studies were obtained either from forearm skin biopsies or from occipital scalp biopsies, designated as non-human subject research under 45 CFR Part 46 and therefore IRB exempted for research.

Whole-exome sequencing

Full exome sequencing, bioinformatics analysis (consisting of initial data processing, base calls, alignments, variant calls, nucleotide and amino acid conservation, biochemical nature of amino acid substitution, population frequency, and predicted functional impact), and filtering based on autosomal and X-linked dominant and recessive inheritance models was performed on the autosomal recessive CGHT case using genomic DNA from the proband (II-1), the unaffected mother (I-2), and the unaffected father (I-1) through Ambry Genetics (Allso Viejo, CA). Evaluation of relationships between the proband and unaffected parents was performed using Short Tandem Repeat markers and samples were prepared using the SureSelect Target Enrichment System (Agilent Technologies). In brief, genomic DNA was sheared, adaptor ligated, and PCR amplified, followed by incubation with the exome baits, elution, and then PCR amplification. Libraries were generated, quantified, and hybridized to the Illumina HiSeq 2000 flow cell for paired-end sequencing. Bioinformatics analysis was used for initial data processing, base calls, alignments as well as variant calls, and the Ambry Genetics Variant Analyzer tool (AVA) was used to determine conservation of nucleotide and amino acids as well as biochemical nature of substitutions, population frequency, and predicted functional impact. Various genome databases (Human Genome Mutation Database (HGMD), HapMap data, Single Nucleotide Polymorphism database (dbSNP), 1000 genomes) were used to search for previously described mutations and/or polymorphisms, and co-segregation studies were performed for candidate gene mutations. Full-exome sequencing revealed a total of 53 genes (85 alterations), and filtering based on the following criteria was used: deleterious nature of the alteration (34 genes, 5 unique alterations), removal of

alterations clearly unrelated to patient's phenotype (26 genes, 33 alterations), and further interpretive analysis based on literature searches for genotype-phenotype correlation resulted in the selection of 3 genes (3 alterations) for further investigation.

Nucleic acid extraction and sequencing

Genomic DNA was isolated from whole blood of the proband (I-1) and father (II-1) from the autosomal recessive CGHT case as well as from the patient with sporadic CGHT. Total RNA was extracted from forearm skin biopsies as well as from cultured keratinocytes and fibroblasts from family members I-1 and II-1 using the Qiagen RNeasy RNA extraction kit and standard methods.

PCR amplification of the exon 32-intron 32 boundary was performed in the autosomal recessive CGHT case using standard conditions and the following primers: F: 5′-GAACATCTTCAGAAGACTGTAAAG-3′ and R: 5′- GTAATCTGAGGATTCCCTAGCATAC -3′. To amplify and sequence the mutant *ABCA5* transcript, the following primers were used: F (in Exon 28): 5′-GCTGATGGGTTGCCAGTGTTGTGAAG-3′, and R (in Exon 33): 5′-CACATGTGCTGTTTGGCTTTGGGATC-3′. For sequencing of *ABCA5* in the sporadic CGHT case, 100–200 ng gDNA was used and primers were designed to flank the intron-exon boundaries 100–150 bp from each exon.

Quantitative PCR on genomic DNA and quantitative RT-PCR on RNA

Quantitative PCR was performed on genomic DNA from the sporadic CGHT case to confirm the 1.3 Mb deletion identified; three amplicons on chromosome 17 were tested using Relative Quantification and the following primers: Amplicon 1 (chr17: 66639201-66639359, 159 bp): F: 5′-TAGATCATTCTCCTAAATGCTCTTCC -3′, R: 5′-GATGCAGCAAAGTTCTCAGGTG -3′; Amplicon 2 (chr17: 66958719- 66958932, 214 bp): F: 5′- GCTGAGCCTCTCCTGAAAACTGGACAAC -3′, R: 5′- GACTCAACTGACATAGGCCATGACAG -3′; Amplicon 3 (chr17: 67249799-67249985, 213 bp): F: 5′- GAACATCTTCAGAAGACTGTAAAG -3′, R: 5′- CTGAGGATTCCCTAGCATACTTAGAGC -3′. Amplicon 4 (chr17: 67716661- 67716904, 244 bp): F: 5′- ACCATGTAAACAAGGAAAACAAC -3′, R: 5′- CTGAGGATTCCCTAGCATACTTAGAGC -3′. Amplicon 5 (chr17: 68403395-68403416, 244 bp): F: 5′- CATTTATCCATATGGGAGGTAG -3′, R: 5′- AACAGATGTCCAAGAGAGTCAAATC -3′. Values were normalized to the *β2M* amplicon (161 bp) using the following primers: F: 5′- CACCTATCCCTGTTGTATTTTATCG -3′, R: 5′-CTCTTTATTTCTGCTGAGGTTTT-3′. All coordinates reference UCSC human reference genome build hg19.

Quantitative RT-PCR was performed on RNA isolated from whole skin, cultured keratinocytes and fibroblasts as previously described [5], [10] and as per the manufacturer's instructions. Relative quantification using the ddCT method [56] was performed with the *β2M* gene as the housekeeping control. The following primers were used for qRT-PCR assays: *ABCA5:* F: 5′-GAACCAACTTCAGGCCAGGTATT-3′, R: 5′-CACATGTGCTGTTTGGCTTTGGGATC -3′; *β2M:* F: 5′-GAGGCTATCCAGCGTACTCCA -3′, R: 5′- CGGCAGGCATACTCATCTTTT-3′; *SOX9:* F: 5′- AGTACCCGCACTTGCACAA-3′, R: 5′-CCGTTCTTCACCGACTTCCT-3′. *GAPDH:* F: 5′-GGAGCGAGATCCCTCCAAAAT -3′, R: 5′- GGCTGTTGTCATACTTCTCATGG-3′

Statistical analysis

All experiments were performed in triplicates using materials from affected, carrier and control individuals. Experiments were repeated in triplicates, and a Student unpaired t-test was used to determine statistical significance in qPCR and qRT-PCR experiments with a p value of 0.05 as the cutoff value for significance. For autophagy experiments, quantification statistical analysis was performed using a two-tailed, equal variance Student's *t*-test. *P*-values of <0.05 (*), <0.01 (**), <0.001 (***) were determined to be statistically significant. The number of cells analyzed per condition is as follows: control, no bafilomycin: 61 cells; control, bafilomycin: 63 cells; affected, no bafilomycin: 37 cells; control, bafilomycin: 43 cells.

Isolation and culturing of human keratinocytes and fibroblasts from whole-skin biopsies

Keratinocytes and fibroblasts were isolated and cultured from whole skin biopsies for the proband (I-1) and unaffected father (II-1) from the first case, and the patient from the second sporadic CGHT case using the method described in [10].

In situ hybridization and immunohistochemical staining on whole skin

The DIG labeling system (Roche) was used to construct the sense and antisense riboprobes for *hABCA5* sequences, amplified using the following primers: *hABCA5* (Exon 13–19, 812 bp): F: 5′-GTGCAGAAGGTTTTACTAGATTTAGACA-3′, R: 5′- GTCTGGAACAAGTTTGATGGGAACCAC-3′; *hABCA5* (Exon 28–33, 550 bp): F: 5′-GCTGATGGGTTGCCAGTGTTGTGAAG-3′, R: 5′-CACATGTGCTGTTTGGCTTTGGGATC -3′. For human expression studies, *in situ* hybridization was performed on 10 μM skin and hair follicle sections from control, carrier and affected individuals using the methods described in [10].

Immunohistochemistry was performed on formalin-fixed paraffin-embedded (FFPE) sections of whole skin and hair follicles. In brief, slides were deparaffinized and rehydrated with a series of ethanol washes and then with 1X TBS. Antigen retrieval was performed for 10 minutes in a 1M sodium citrate, pH 6.0 solution heated to 95°C, and then slides were cooled and washed three times in 1X TBS. Tissues were blocked in either 10% normal donkey serum (Jackson ImmunoResearch, PA, USA) or 2% fish skin gelatin (Sigma Aldrich, MO, USA) and incubated with primary antibody in 1X TBS at 4°C overnight. The anti -rabbit ABCA5 antibody (ab99953, Abcam) was used at a concentration of 1:200 and the anti -rabbit IgG isotype primary antibody (Santa Cruz Biotechnologies, CA, USA) control was used at the same concentration. Slides were then washed with 1X TBS, incubated with the goat anti -rabbit biotin-conjugated antibody (1:800 in 1X TBS) for an hour at room temperature, washed again with 1X TBS and then incubated with the streptavidin-alkaline phosphatase (AP) tertiary antibody (Invitrogen; 1:300 in 1X TBS) for 30 minutes at room temperature. The SIGMA *FAST* Fast Red TR/Naphthol AS-MX Tablets (Sigma Aldrich, MO, USA) were used to develop the slides, which were then mounted with Dako Glycergel Mounting Medium (Dako, CA, USA).

Immunofluorescence staining and immunoblotting

Immunofluorescence staining was performed on whole skin 10 μm sections from both human and mouse embedded in Optimal Cutting Temperature (O.C.T.). Slides were fixed with 50% MeOH/50% Acetone for 10 minutes at −20°C, washed with 1X PBS, and then blocked with 2% fish skin gelatin (Sigma Aldrich, MO, USA). The ABCA5 (Abcam, ab99953) and anti-rabbit IgG (Santa Cruz Biotechnologies) primary antibodies were used at a concentration of 1:200. Slides were then washed,

incubated with the Alexa Fluor 488 donkey anti-rabbit IgG (Molecular Probes, Invitrogen) secondary antibody (1:800 in 1X PBS), mounted with VECTASHIELD mounting medium with DAPI (Vector Laboratories, Burlingame, CA, USA), and imaged using a LSM 5 laser-scanning Axio Observer Z1 confocal microscope (Carl Zeiss).

For autophagy analysis, human keratinocytes were grown to confluence, seeded on 12 mm coverslips, and then fixed with 4% paraformaldehyde for 20 min at room temperature. After permeabilization with 200 μg/ml digitonin (Invitrogen) in PBS for 10 min, cells were incubated with the specified primary antibodies for 1 hr at room temperature. Subsequently, cells were incubated with the appropriate Alexa-Fluor-conjugated secondary antibodies for 1 hr at room temperature, for cholesterol staining cells were additionally incubated with Filipin complex (Sigma) for 1 hr at room temperature. Images were acquired by confocal laser scanning microscopy (Zeiss LSM-700) and analyzed with Zeiss Zen and ImageJ Software (NIH). The number of LC3-positive compartments and their surface areas (expressed as number of pixels per field) were normalized to the number of cells in each field. The average size was obtained by dividing the surface area of the LC3-positive compartment (in pixel^2) by the number of LC3 puncta. Similar measurements were made for Lamp1 and p62 compartments. Primary antibodies used for immunofluorescence: mouse anti-LC3 (MBL), guinea pig anti-p62 (Progen), rabbit anti-Lamp1 (Abcam) mouse anti-Lamp2 (Santa Cruz). Bafilomycin A1 (50 nM, Wako) was added to the media 2 hours prior to cell fixation.

Immunoblotting was performed using 10 μg total protein extracted with RIPA buffer/proteinase inhibitor cocktail from cultured keratinocytes and fibroblasts and SDS-PAGE was used to separate proteins, followed by a wet transfer to a hybond ECL nitrocellulose (Amersham, NJ, USA) or PVDF membrane (BioRad, CA, USA) membrane. All membranes were blocked with 5% milk for one hour at room temperature and then incubated with the primary antibodies, ABCA5 (1:500) and beta-actin (1:1000; Santa Cruz Biotechnologies, CA, USA) diluted in Washing Buffer (1X PBS 0.1% Tween-20). Membranes were then washed several times in 15-minute intervals with 1X PBST, incubated with goat anti-rabbit or mouse- HRP conjugated secondary antibody (Invitrogen; 1:1000), and then developed with the SuperSignal West Dura Extended Duration Substrate (Thermo Scientific, IL, USA).

Cytogenetic analyses

G-banding analysis was performed using standard techniques, and FISH using chromosomal paint was performed on metaphase chromosomes obtained from peripheral blood leukocytes (PBLs), as per the manufacturer's directions (VYSIS). Sub-telomeric FISH was performed using a mixture of probes for chromosomes 17q, 17centromere, 9p, and 9q obtained from VYSIS in accordance with the manufacturer's instructions. Chromosomes were counterstained with DAPI (VYSIS) and hybridized metaphase chromosomes were viewed using a Nikon microscope fitted with a filter wheel and Cytovision Applied Imaging software.

Breakpoint mapping and copy number variant (CNV) analysis

Breakpoint mapping was performed using FISH with BAC clones in the chromosome 17q24 region. To fine map the deleted region on chromosome 17q, genome scanning was initially performed using the Affymetrix 500k whole-genome mapping array, and then the 2.7M array (Affymetrix) to further examine CNVs in the region of the breakpoint.

Supporting Information

Figure S1 Histological analysis of CGHT and control hair follicles by hematoxylin and eosin staining. (A–C) Hematoxylin and eosin staining of a patient skin biopsy from the forearm demonstrated that hair follicles are of the terminal type, as they are medullated, pigmented, and penetrate deep in the dermis. (B) Enlarged image of a late catagen hair follicle. (C) Magnified image of an anagen hair follicle situated deep in the dermis (A). Note the thickness of the outer root sheath compared to control hair follicles (D–F). (D) Hematoxylin and eosin staining of a control skin biopsy from the forearm. (E) Magnified image of a hair in catagen. (F) Enlarged image of the apoptosing strand of the catagen hair follicle. Note the size of the control hair follicle compared to patient hair follicles. No anagen hair follicles were present in the control skin biopsy. DS = dermal sheath; PFD = perifollicular dermis; IRS = inner root sheath; ORS = outer root sheath; HS = hair shaft; SG = sebaceous gland; DP = dermal papilla.
(TIF)

Figure S2 Mouse Abca5 localization pattern in the adult testis and epididymis by immunohistochemistry and immunofluorescence staining. (A, B) Immunohistochemistry on formalin-fixed paraffin-embedded (FFPE) adult mouse testis demonstrated strong localization of Abca5 to the basal cells of the seminiferous tubules (arrows in (A)), interstitial space consisting of Leydig cells, and tunica albuginea (arrow in (B)). (D, E) Abca5 immunohistochemistry on formalin-fixed paraffin-embedded adult mouse epididymis revealed strong localization to the connective tissue, smooth muscle cells and fibrocytes surrounding the cylindrical epithelium within the corpus and cauda regions, as well as within the basal and tall columnar cells of the cauda cylindrical epithelium (E). (C, F) Testis and epididymis sections incubated without primary antibody produced no signal. (G–I) Immunofluorescence staining of Abca5 in the testis and epididymis demonstrated Abca5 localization to the same structures within the testis and epididymis as did the immunohistochemical staining method.
(TIF)

Figure S3 Mouse Abca5 localizes to the outer and inner root sheath of anagen hair follicles. (A, B) Immunofluorescence staining of Abca5 on frozen mouse anagen skin (day 30) sections demonstrated a signal only within hair follicles (A), specifically within the outer and inner root sheath (B, C). (E–F) Immunohistochemical staining on formalin-fixed paraffin-embedded (FFPE) anagen skin sections revealed Abca5 localization to the outer and inner root sheath, and some signal present in the follicular dermis including the dermal sheath. No signal was observed on sections incubated without primary antibody (D, G).
(TIF)

Figure S4 Immunoblotting on protein extracted from carrier and patient fibroblasts in the presence of absence of the enzyme, PNGaseF that removes all *N*-glycosyl modifications revealed loss of a ~100 kDa band that is the glycosylated form of the protein in the patient relative to the carrier. β-actin was used as a loading control.
(TIF)

Figure S5 Quantification of immunofluorescence staining for LC3 and p62 reveals defective autophagic clearance in CGHT. (A) The formation of LC3 puncta was significantly increased in affected vs. control keratinocytes ($p<0.05$, – Bafilomycin; $p<0.01$, + Bafilomycin) as well as within control keratinocytes + Bafilomycin vs. – Bafilomycin treatment ($p<0.01$), but no significant difference was observed between affected keratinocytes + Bafilomycin vs. – Bafilomycin treatment. (B) The formation of

p62 puncta was significantly increased in affected vs. control keratinocytes ($p<0.01$) as well as within control keratinocytes + Bafilomycin vs. – Bafilomycin treatment ($p<0.01$), but no significant difference was observed between affected keratinocytes + Bafilomycin vs. – Bafilomycin treatment. A Student t test (unpaired) was performed with a cutoff *P* value of 0.05 for statistical significance and error bars represent the standard deviation. ImageJ was used for image quantification.
(TIF)

Figure S6 Histological analysis of hair follicles from the sporadic CGHT case**.** Hematoxylin and eosin staining of a patient skin biopsy from the lower back reveals the presence of terminal hair follicles in the anagen stage. DS = dermal sheath; PFD = perifollicular dermis; IRS = inner root sheath; ORS = outer root sheath; HC = hair canal; SG = sebaceous gland.
(TIF)

Figure S7 Telomere FISH and FISH using BAC clones spanning chromosome 17q24.2-24.3 to detect the 1.3 Mb cryptic deletion in sporadic CGHT**.** (A) Telomere FISH was performed to test possible deletions at the end of chromosome 17q using a commercially available probe mix for 17q (green), 17 centromere (green), 9p (green), and 9q (red). Note the presence of the green signal on the derived chromosome 3 indicating that the telomere of chromosome 17 was not deleted in the t3;17 rearrangement. (B–C) FISH using BAC clones on chromosome 17q24.2-q24.3 revealed a cryptic deletion at the breakpoint of chromosome 17. Metaphase spreads and interphase nuclei show only one signal for clones RP11-387O17 (green) (B) and RP11-293K20 (green) (C), which hybridize to the deleted 1.3 Mb portion of chromosome 17q24.2-24.3.
(TIF)

Figure S8 Summary of CNVs within the chr17q24.2-24.3 region identified in autosomal dominant and sporadic cases of CGHT illustrates that ABCA5 is located in the minimal common region. *ABCA5* (red box) and the other genes in the surrounding region (black boxes) as well as direction of transcription (arrows) are indicated. Nature of the variants (duplications, deletions) previously reported as well as identified in this study is indicated as well as the sizes and corresponding references. qPCR amplicons are represented by vertical lines, where two amplicons flank the 1.3 Mb deleted region (red box) and three amplicons lie within it. Database of Genomic Variants (DGV) alterations are indicated as gray boxes. All variants lie 1–2 Mb upstream of the *SOX9* gene. All coordinates reference the UCSC Genome Browser human reference genome hg19.
(TIF)

Acknowledgments

We would like to thank the patients their families for participating in this study. We are very grateful to Mr. Ming Zhang for his exceptional technical assistance, Drs. Zongyou Guo and Karl Gledhill for cell culture assistance, Dr. Yutaka Shimomura (Niigata University, Japan) and the Hayden lab (Columbia University) for helpful suggestions and trouble-shooting advice on immunoblotting experiments, and members of the A.M.C. laboratory for their insightful discussions. We appreciate the support and expert advice from the Skin Disease Research Center in the Department of Dermatology at Columbia University Medical Center.

Author Contributions

Conceived and designed the experiments: GMD AMC. Performed the experiments: GMD. Analyzed the data: GMD CD GDP. Contributed reagents/materials/analysis tools: CD GDP. Wrote the paper: GMD. Performed autophagy analysis and filipin staining: CD. Performed experiments and analyses in sporadic CGHT case: GMD MKu VJ KAF MKi MTS AS AV DW BL. Performed clinical evaluations of the patients and provided clinical photos: KAY HF NS LR LDLC JCSA.

References

1. Beighton P (1970) Congenital hypertrichosis lanuginosa. Arch Dermatol 101: 669–672.
2. Wendelin DS, Pope DN, Mallory SB (2003) Hypertrichosis. J Am Acad Dermatol 48: 161–179; quiz 180–161.
3. Garcia-Cruz D, Figuera LE, Cantu JM (2002) Inherited hypertrichoses. Clin Genet 61: 321–329.
4. Fantauzzo KA, Kurban M., Levy B., Christiano AM. (2012) Trps1 and its Target Gene Sox9 Regulate Epithelial Proliferation in the Developing Hair Follicle and are Associated with Hypertrichosis. PLoS Genet 8: e1003002.
5. Fantauzzo KA, Tadin-Strapps M, You Y, Mentzer SE, Baumeister FA, et al. (2008) A position effect on TRPS1 is associated with Ambras syndrome in humans and the Koala phenotype in mice. Hum Mol Genet 17: 3539–3551.
6. Sun M, Li N, Dong W, Chen Z, Liu Q, et al. (2009) Copy-number mutations on chromosome 17q24.2-q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia. Am J Hum Genet 84: 807–813.
7. Tadin M, Braverman E, Cianfarani S, Sobrino AJ, Levy B, et al. (2001) Complex cytogenetic rearrangement of chromosome 8q in a case of Ambras syndrome. Am J Med Genet 102: 100–104.
8. Tadin-Strapps M, Salas-Alanis JC, Moreno L, Warburton D, Martinez-Mir A, et al. (2003) Congenital universal hypertrichosis with deafness and dental anomalies inherited as an X-linked trait. Clin Genet 63: 418–422.
9. Zhu H, Shang D, Sun M, Choi S, Liu Q, et al. (2011) X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. Am J Hum Genet 88: 819–826.
10. DeStefano GM, Fantauzzo KA, Petukhova L, Kurban M, Tadin-Strapps M, et al. (2013) Position effect on FGF13 associated with X-linked congenital generalized hypertrichosis. Proc Natl Acad Sci U S A 110: 7790–7795.
11. van Blitterswijk WJ, Houssa B (2000) Properties and functions of diacylglycerol kinases. Cell Signal 12: 595–605.
12. Bordi F, Ugolini A (1999) Group I metabotropic glutamate receptors: implications for brain diseases. Prog Neurobiol 59: 55–79.
13. Pacheco MA, Jope RS (1996) Phosphoinositide signaling in human brain. Prog Neurobiol 50: 255–273.
14. Rodriguez de Turco EB, Tang W, Topham MK, Sakane F, Marcheselli VL, et al. (2001) Diacylglycerol kinase epsilon regulates seizure susceptibility and long-term potentiation through arachidonoyl- inositol lipid signaling. Proc Natl Acad Sci U S A 98: 4740–4745.
15. Kubo Y, Sekiya S, Ohigashi M, Takenaka C, Tamura K, et al. (2005) ABCA5 resides in lysosomes, and ABCA5 knockout mice develop lysosomal disease-like symptoms. Mol Cell Biol 25: 4138–4149.
16. Ye D, Meurs I, Ohigashi M, Calpe-Berdiel L, Habets KL, et al. (2010) Macrophage ABCA5 deficiency influences cellular cholesterol efflux and increases susceptibility to atherosclerosis in female LDLr knockout mice. Biochem Biophys Res Commun 395: 387–394.
17. Reese MG, Eeckman FH, Kulp D, Haussler D (1997) Improved splice site detection in Genie. J Comput Biol 4: 311–323.
18. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, et al. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet 15: 2490–2508.
19. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR (2003) ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31: 3568–3571.
20. Petry F, Kotthaus A, Hirsch-Ernst KI (2003) Cloning of human and rat ABCA5/Abca5 and detection of a human splice variant. Biochem Biophys Res Commun 300: 343–350.
21. Petry F, Ritz V, Meineke C, Middel P, Kietzmann T, et al. (2006) Subcellular localization of rat Abca5, a rat ATP-binding-cassette transporter expressed in Leydig cells, and characterization of its splice variant apparently encoding a half-transporter. Biochem J 393: 79–87.
22. Draheim V, Reichel A, Weitschies W, Moenning U (2010) N-glycosylation of ABC transporters is associated with functional activity in sandwich-cultured rat hepatocytes. Eur J Pharm Sci 41: 201–209.
23. Benavides F, Starost MF, Flores M, Gimenez-Conti IB, Guenet JL, et al. (2002) Impaired hair follicle morphogenesis and cycling with abnormal epidermal differentiation in nackt mice, a cathepsin L-deficient mutation. Am J Pathol 161: 693–703.
24. Tobin DJ, Foitzik K, Reinheckel T, Mecklenburg L, Botchkarev VA, et al. (2002) The lysosomal protease cathepsin L is an important regulator of

keratinocyte and melanocyte differentiation during hair follicle morphogenesis and cycling. Am J Pathol 160: 1807–1821.
25. Roth W, Deussing J, Botchkarev VA, Pauly-Evers M, Saftig P, et al. (2000) Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and pertubation of hair follicle cycling. FASEB J 14: 2075–2086.
26. Mannan AU, Roussa E, Kraus C, Rickmann M, Maenner J, et al. (2004) Mutation in the gene encoding lysosomal acid phosphatase (Acp2) causes cerebellum and skin malformation in mouse. Neurogenetics 5: 229–238.
27. Dall'Armi C, Devereaux KA, Di Paolo G (2013) The role of lipids in the control of autophagy. Curr Biol 23: R33–45.
28. Bornig H, Geyer G (1974) Staining of cholesterol with the fluorescent antibiotic "filipin". Acta Histochem 50: 110–115.
29. Vanier MT (2010) Niemann-Pick disease type C. Orphanet J Rare Dis 5: 16.
30. Nowak JA, Polak L, Pasolli HA, Fuchs E (2008) Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell 3: 33–43.
31. Vidal VP, Chaboissier MC, Lutzkendorf S, Cotsarelis G, Mill P, et al. (2005) Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Curr Biol 15: 1340–1351.
32. Shimomura Y, Garzon MC, Kristal L, Shapiro L, Christiano AM (2009) Autosomal recessive woolly hair with hypotrichosis caused by a novel homozygous mutation in the P2RY5 gene. Exp Dermatol 18: 218–221.
33. Shimomura Y, Wajid M, Ishii Y, Shapiro L, Petukhova L, et al. (2008) Disruption of P2RY5, an orphan G protein-coupled receptor, underlies autosomal recessive woolly hair. Nat Genet 40: 335–339.
34. Shimomura Y, Wajid M, Petukhova L, Shapiro L, Christiano AM (2009) Mutations in the lipase H gene underlie autosomal recessive woolly hair/hypotrichosis. J Invest Dermatol 129: 622–628.
35. Chi W, Wu E, Morgan BA (2013) Dermal papilla cell number specifies hair size, shape and cycling and its reduction causes follicular decline. Development 140: 1676–1683.
36. Sharov AA, Sharova TY, Mardaryev AN, Tommasi di Vignano A, Atoyan R, et al. (2006) Bone morphogenetic protein signaling regulates the size of hair follicles and modulates the expression of cell cycle-associated genes. Proc Natl Acad Sci U S A 103: 18166–18171.
37. Ohtsuki S, Kamoi M, Watanabe Y, Suzuki H, Hori S, et al. (2007) Correlation of induction of ATP binding cassette transporter A5 (ABCA5) and ABCB1 mRNAs with differentiation state of human colon tumor. Biol Pharm Bull 30: 1144–1146.
38. Heimerl S, Bosserhoff AK, Langmann T, Ecker J, Schmitz G (2007) Mapping ATP-binding cassette transporter gene expression profiles in melanocytes and melanoma cells. Melanoma Res 17: 265–273.
39. Saini V, Hose CD, Monks A, Nagashima K, Han B, et al. (2012) Identification of CBX3 and ABCA5 as putative biomarkers for tumor stem cells in osteosarcoma. PLoS One 7: e41401.
40. Fletcher JI, Haber M, Henderson MJ, Norris MD (2010) ABC transporters in cancer: more than just drug efflux pumps. Nat Rev Cancer 10: 147–156.
41. Lefevre C, Audebert S, Jobard F, Bouadjar B, Lakhdar H, et al. (2003) Mutations in the transporter ABCA12 are associated with lamellar ichthyosis type 2. Hum Mol Genet 12: 2369–2378.
42. Kelsell DP, Norgett EE, Unsworth H, Teh MT, Cullup T, et al. (2005) Mutations in ABCA12 underlie the severe congenital skin disease harlequin ichthyosis. Am J Hum Genet 76: 794–803.
43. Ringpfeil F, Lebwohl MG, Christiano AM, Uitto J (2000) Pseudoxanthoma elasticum: mutations in the MRP6 gene encoding a transmembrane ATP-binding cassette (ABC) transporter. Proc Natl Acad Sci U S A 97: 6001–6006.
44. Harakalova M, van Harssel JJ, Terhal PA, van Lieshout S, Duran K, et al. (2012) Dominant missense mutations in ABCC9 cause Cantu syndrome. Nat Genet 44: 793–796.
45. van Bon BW, Gilissen C, Grange DK, Hennekam RC, Kayserili H, et al. (2012) Cantu syndrome is caused by mutations in ABCC9. Am J Hum Genet 90: 1094–1101.
46. Akiyama M, Sugiyama-Nakagiri Y, Sakai K, McMillan JR, Goto M, et al. (2005) Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer. J Clin Invest 115: 1777–1784.
47. Uitto J, Pulkkinen L, Ringpfeil F (2001) Molecular genetics of pseudoxanthoma elasticum: a metabolic disorder at the environment-genome interface? Trends Mol Med 7: 13–17.
48. Nusse R (2003) Wnts and Hedgehogs: lipid-modified proteins and similarities in signaling mechanisms at the cell surface. Development 130: 5297–5305.
49. Gritli-Linde A, Hallberg K, Harfe BD, Reyahi A, Kannius-Janson M, et al. (2007) Abnormal hair development and apparent follicular transformation to mammary gland in the absence of hedgehog signaling. Dev Cell 12: 99–112.
50. Narhi K, Jarvinen E, Birchmeier W, Taketo MM, Mikkola ML, et al. (2008) Sustained epithelial beta-catenin activity induces precocious hair development but disrupts hair follicle down-growth and hair shaft formation. Development 135: 1019–1028.
51. Parimoo S, Zheng Y, Eilertsen K, Ge L, Prouty S, et al. (1999) Identification of a novel SCD gene and expression of the SCD gene family in mouse skin. J Investig Dermatol Symp Proc 4: 320–322.
52. Zheng Y, Eilertsen KJ, Ge L, Zhang L, Sundberg JP, et al. (1999) Scd1 is expressed in sebaceous glands and is disrupted in the asebia mouse. Nat Genet 23: 268–270.
53. Gates AH, Karasek M (1965) Hereditary Absence of Sebaceous Glands in the Mouse. Science 148: 1471–1473.
54. Panicker SP, Ganguly T, Consolo M, Price V, Mirmirani P, et al. (2012) Sterol intermediates of cholesterol biosynthesis inhibit hair growth and trigger an innate immune response in cicatricial alopecia. PLoS One 7: e38449.
55. Nakamura M, Schneider MR, Schmidt-Ullrich R, Paus R (2012) Mutant laboratory mice with abnormalities in hair follicle morphogenesis, cycling, and/or structure: an update. J Dermatol Sci 69: 6–29.
56. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402–408.

Figure S1

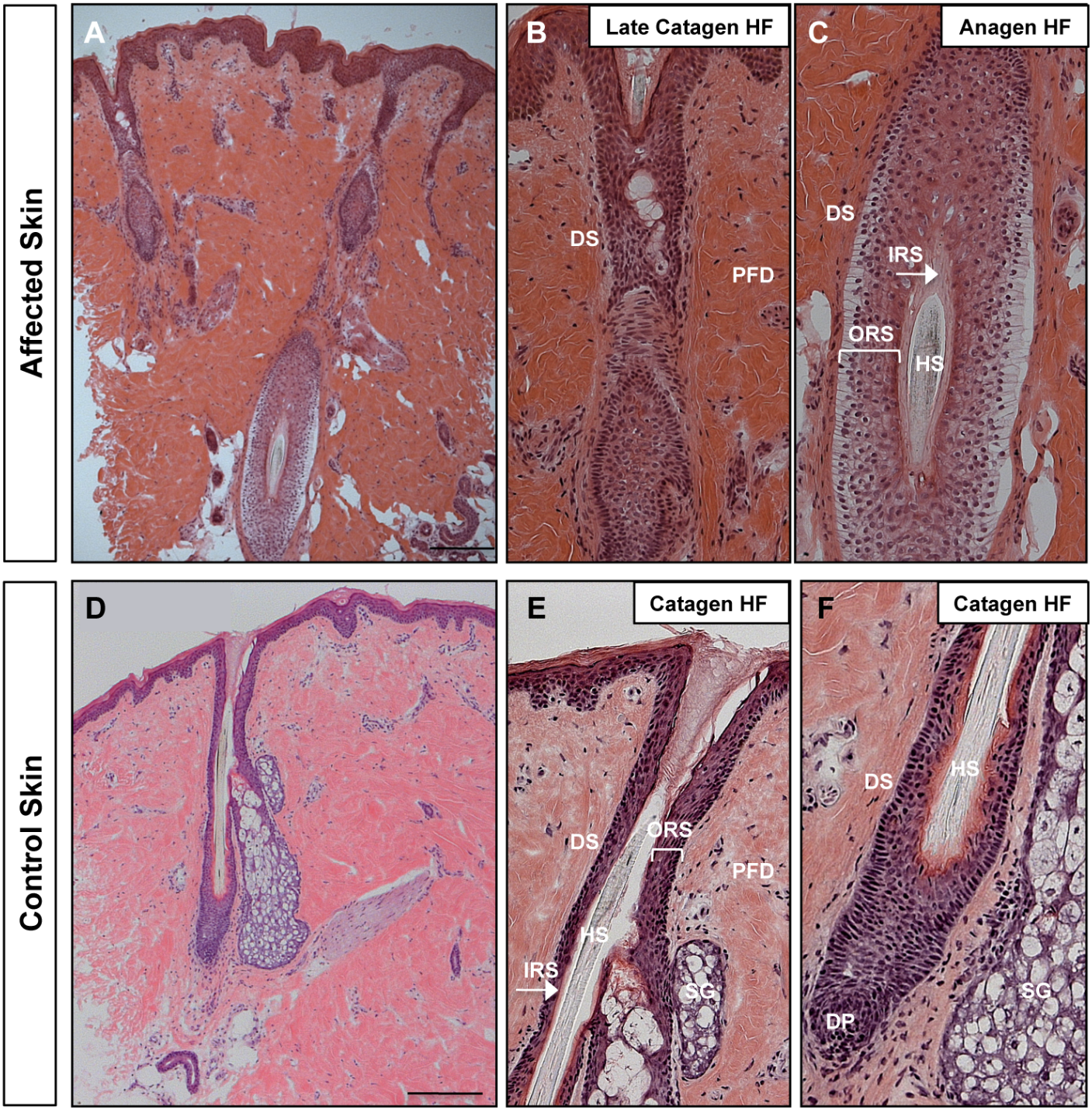

Affected Skin
Control Skin
A
B
Late Catagen HF
DS
PFD
C
Anagen HF
DS
IRS
ORS
HS
D
E
Catagen HF
DS
ORS
PFD
HS
IRS
SG
F
Catagen HF
DS
HS
SG
DP

Figure S2

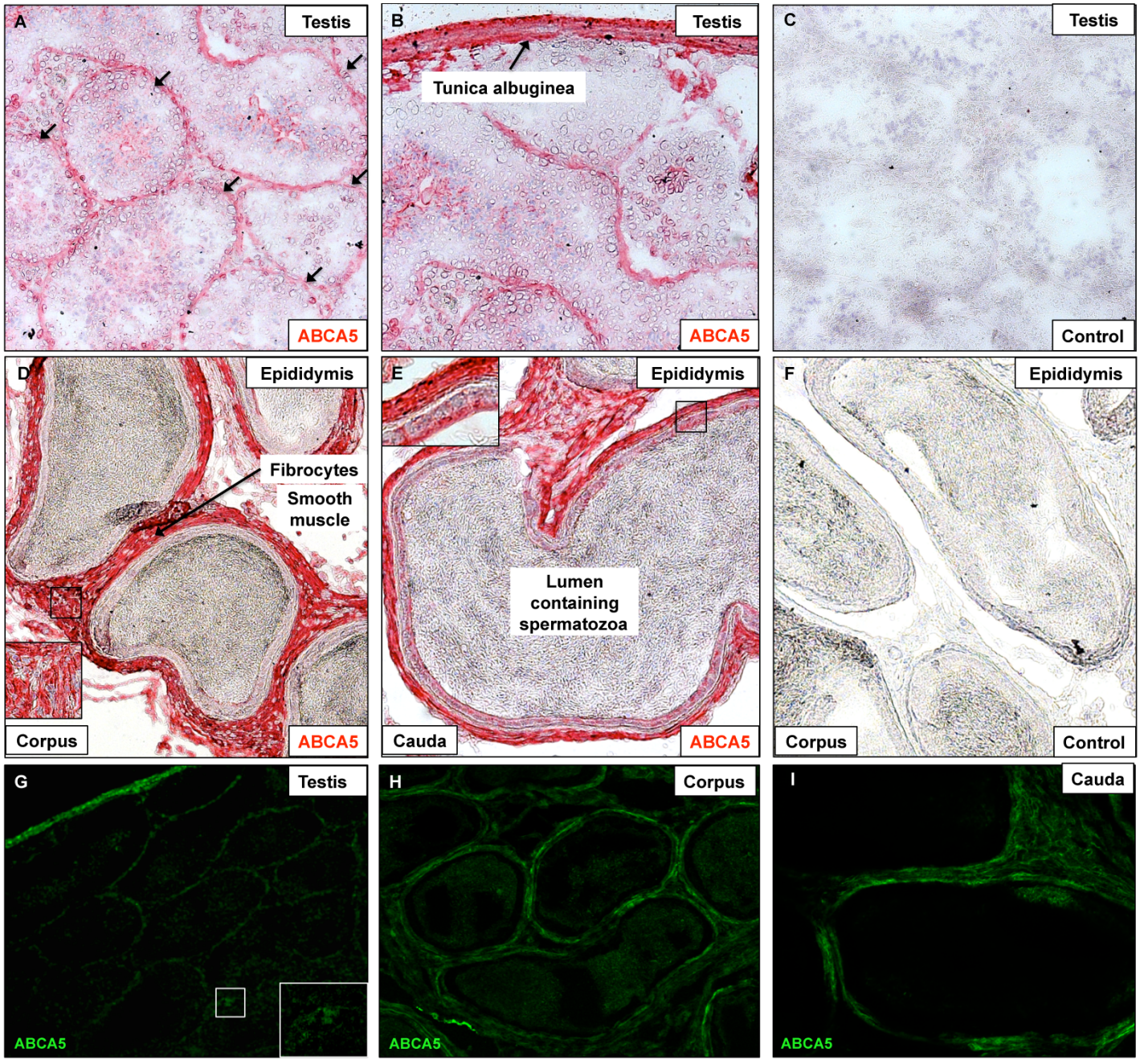

A
Testis
ABCA5
B
Testis
Tunica albuginea
ABCA5
C
Testis
Control
D
Epididymis
Fibrocytes
Smooth muscle
Corpus
ABCA5
E
Epididymis
Lumen containing spermatozoa
Cauda
ABCA5
F
Epididymis
Corpus
Control
G
Testis
ABCA5
H
Corpus
ABCA5
I
Cauda
ABCA5

Figure S3

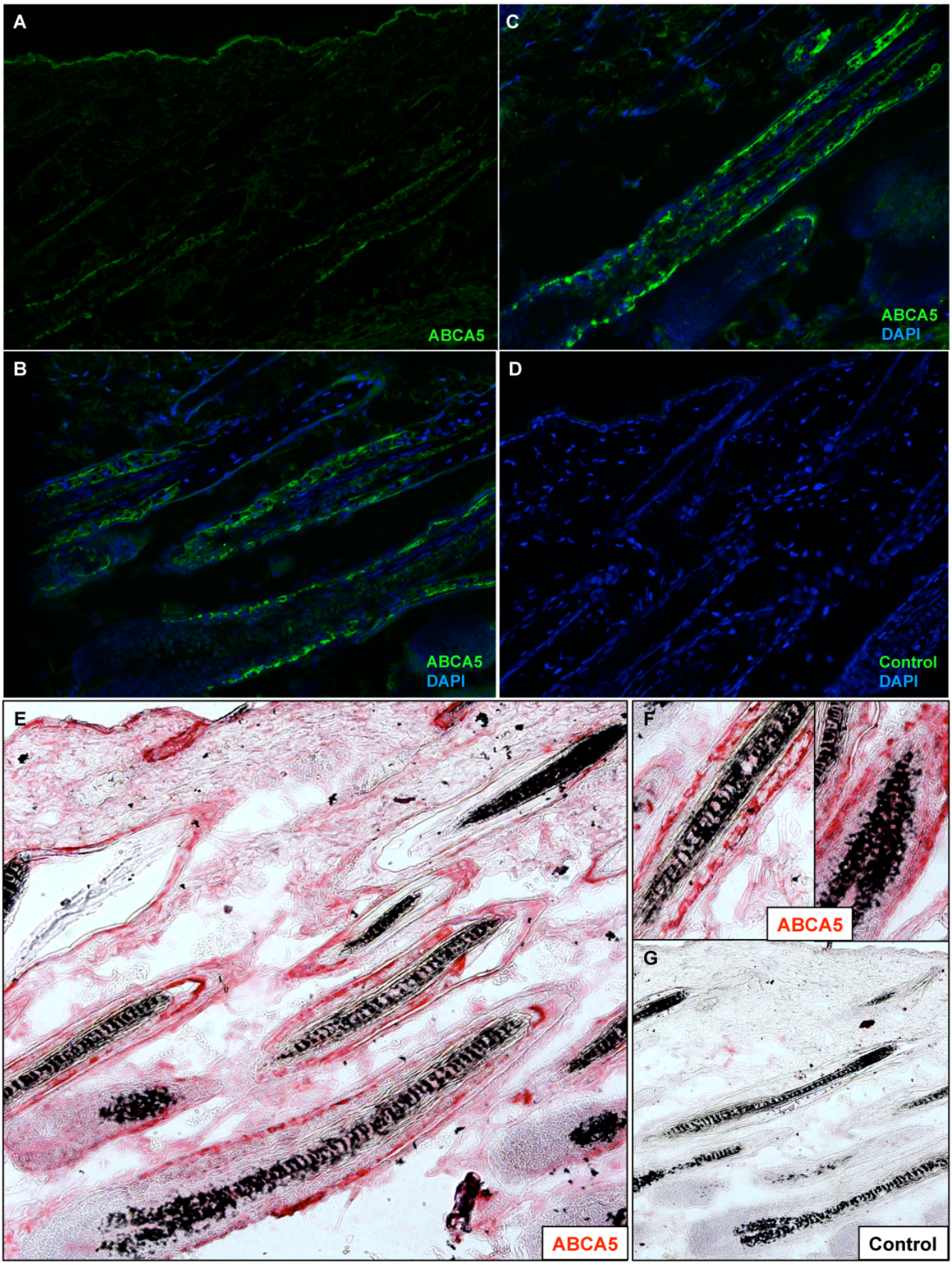

A
ABCA5
C
ABCA5
DAPI
B
ABCA5
DAPI
D
Control
DAPI
E
ABCA5
F
ABCA5
G
Control

Figure S4

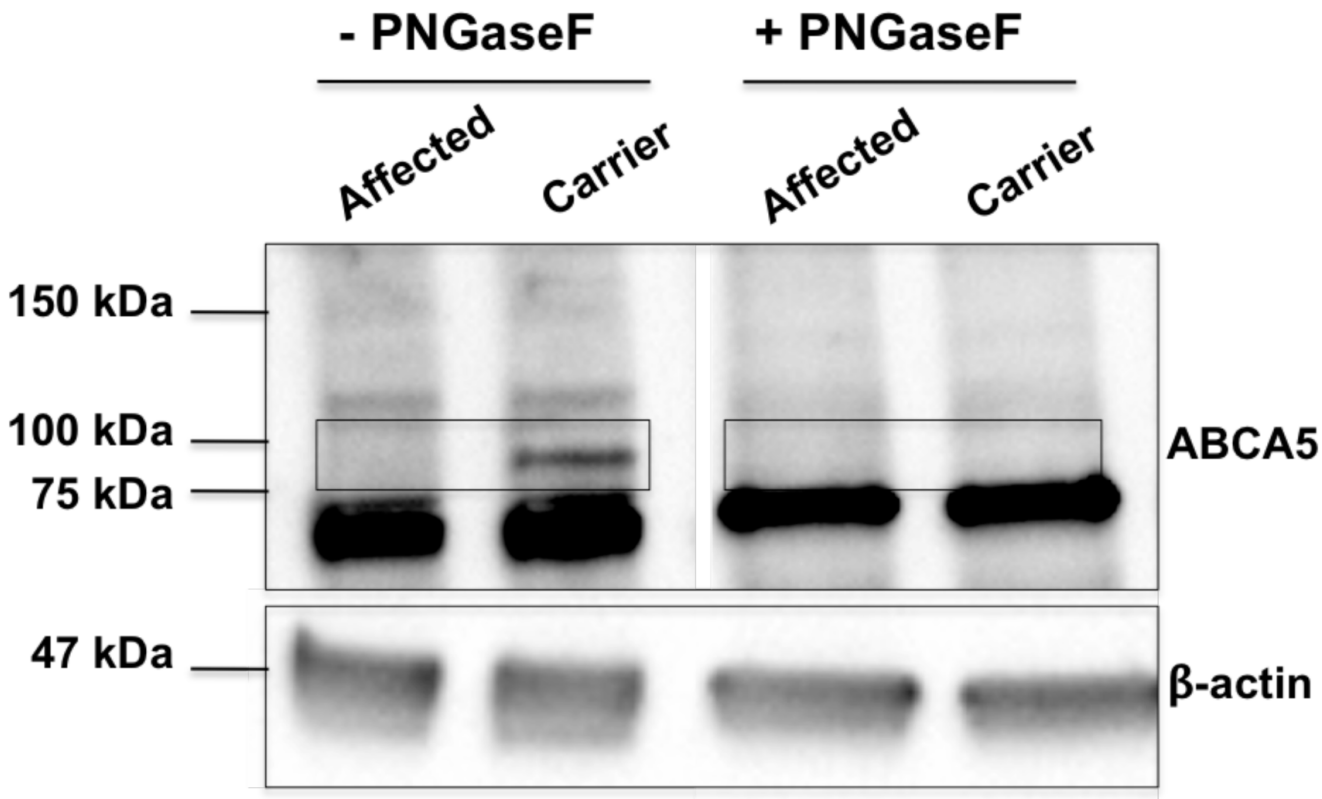


Figure S5

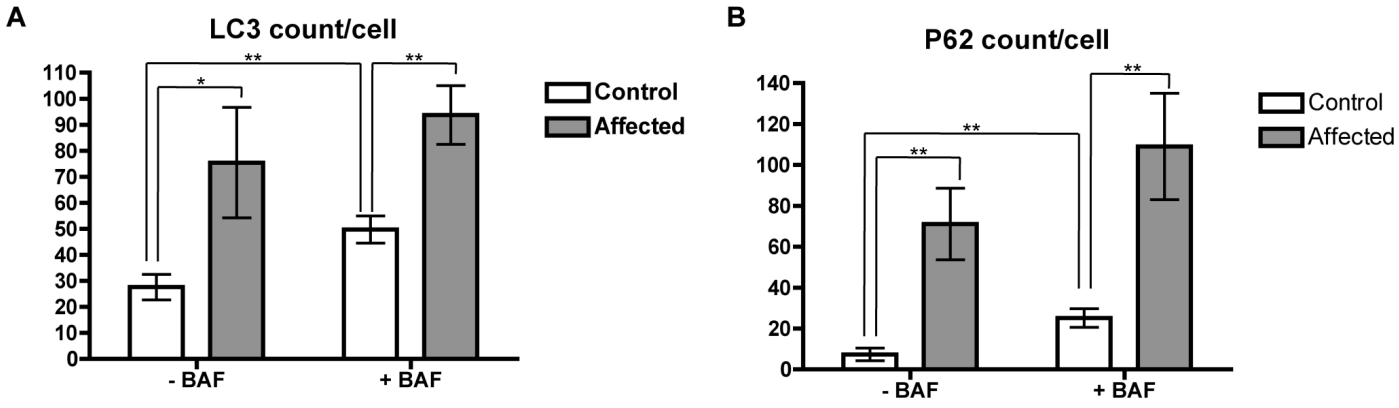


Figure S6

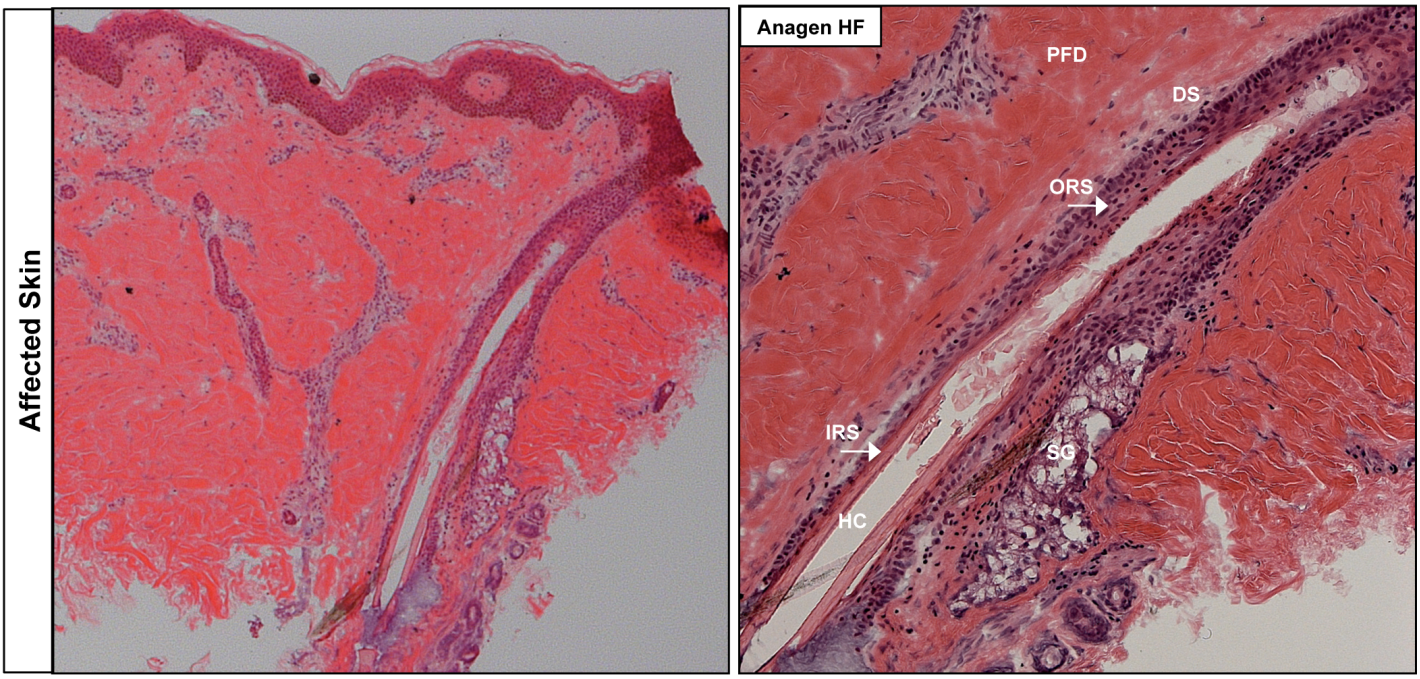


Figure S7

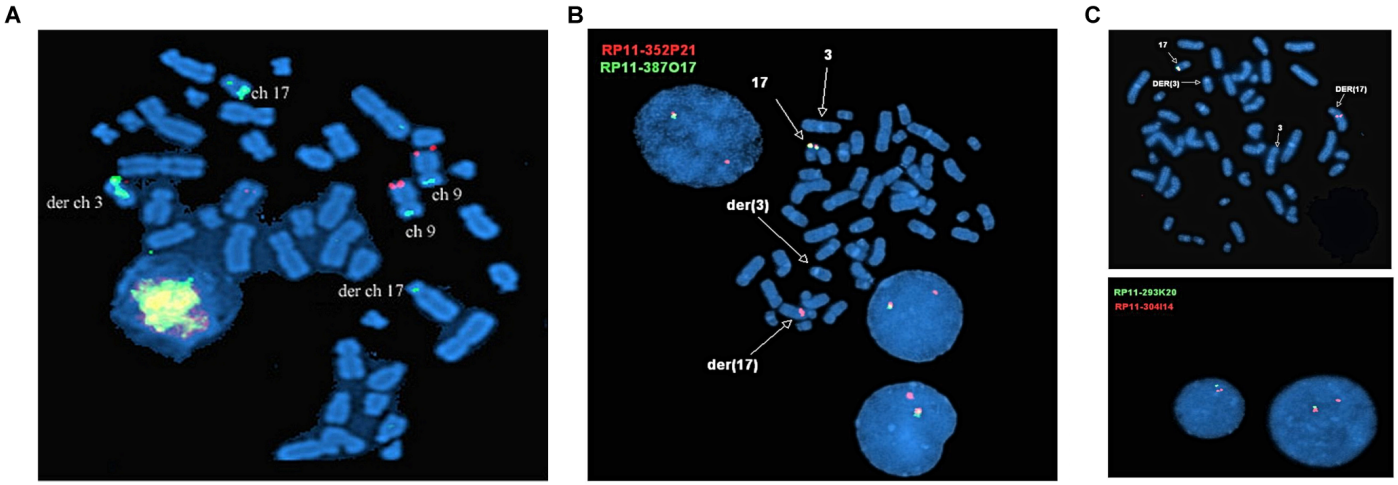
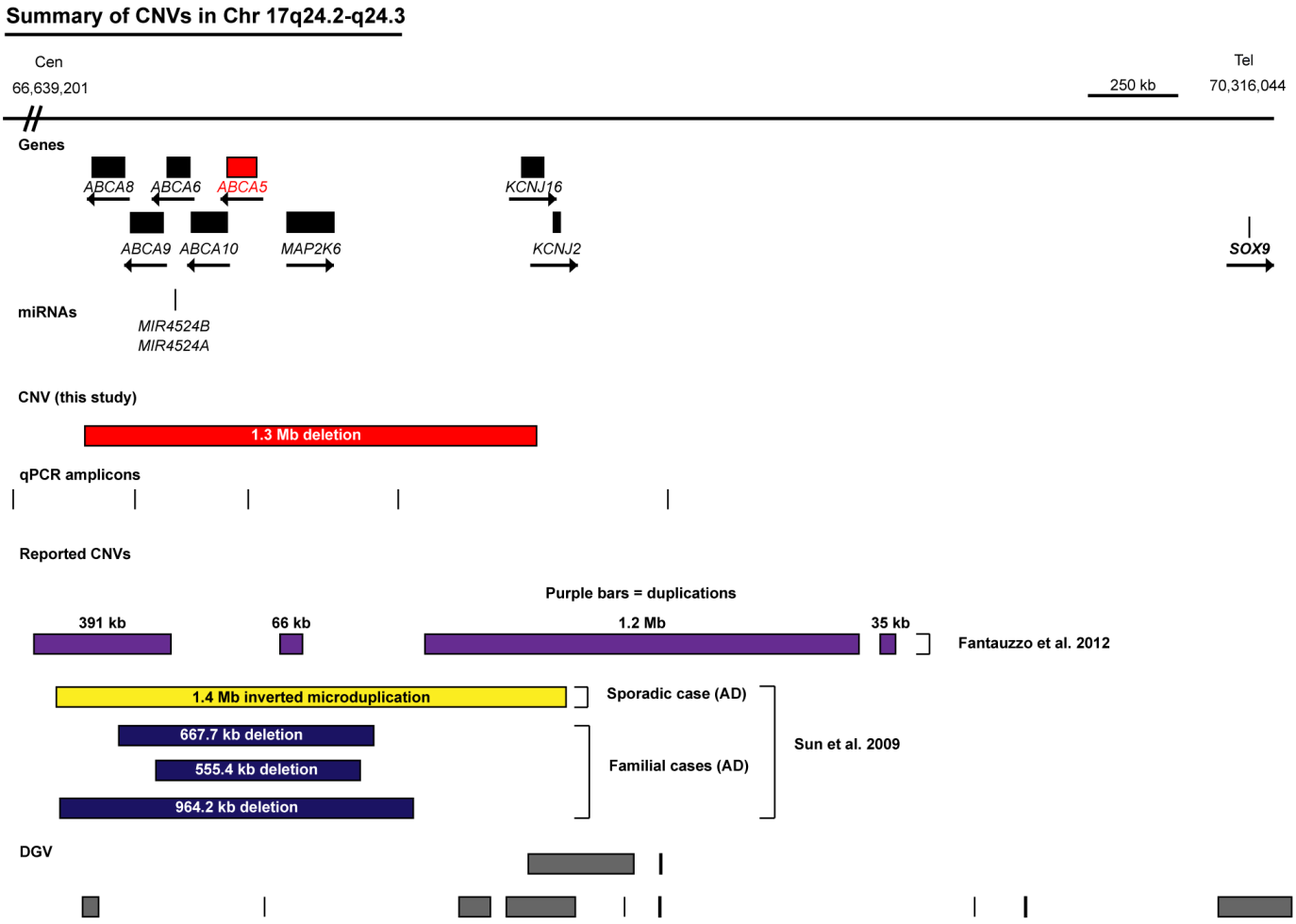


Figure S8


Summary of CNVs in Chr 17q24.2-q24.3
Cen
66,639,201
250 kb
Tel
70,316,044
Genes
ABCA8
ABCA6
ABCA5
KCNJ16
ABCA9
ABCA10
MAP2K6
KCNJ2
SOX9
miRNAs
MIR4524B
MIR4524A
CNV (this study)
1.3 Mb deletion
qPCR amplicons
Reported CNVs
Purple bars = duplications
391 kb
66 kb
1.2 Mb
35 kb
Fantauzzo et al. 2012
1.4 Mb inverted microduplication
Sporadic case (AD)
667.7 kb deletion
555.4 kb deletion
Familial cases (AD)
964.2 kb deletion
Sun et al. 2009
DGV

Chapter III.3

Discussion

Chapter III.3 Discussion

In rare, familial conditions, the genetic architecture is often composed of one or a few mutations that produce large effects at a high penetrance. The majority of such variants can be found in the exome, which represents ~1% of the genome (~30 Mb), and mutations affecting protein coding sequence and splicing are more likely to produce deleterious effects compared to common and non-coding variants. Families containing multiple affected members for a given rare trait with a clear Mendelian inheritance pattern may be sufficiently well-powered for linkage analysis. This method, combined with sequencing (targeted or whole-exome) and genotyping-based studies (i.e. homozygosity mapping) is optimal for narrowing the search for shared pathogenic variants with large effects. However, this approach is not applicable when only a single or a few individuals are affected. Thus, in the absence of positional information and/or a known candidate gene, the ability to sequence the exome is a robust and unbiased approach that has proven highly fruitful for gene discovery in Mendelian disease.

Characterizing the genetic basis of autosomal recessive and sporadic CGHT

In Chapter III.1, I characterized the genetic defect underlying a rare, familial condition affecting hair growth. The approach I used in this study included whole-exome sequencing and filtering for rare variants with large effect sizes, taking into account the autosomal recessive mode of inheritance resulting from consanguinity. To narrow the list of candidate genes, I performed expression

analysis in human skin and hair follicle tissues based on the assumption that the candidate gene would be expressed at the site of pathology for the phenotype, although this does not necessarily hold true for genes that act in a non-cell autonomous manner. Importantly, the key factor in this filtering approach was genomic position and association with reported CNVs associated with CGHT. The identification of a homozygous recessive mutation in a gene previously associated with the disease pathogenesis based on genomic position provided strong evidence for the *ABCA5* gene as a candidate. In the context of function, mutations in the family of ABC transporters have been reported in several inherited conditions manifesting dermatological phenotypes (described in Chapter III.2 and the following section) (QUAZI AND MOLDAY 2011), which provided an even stronger case for the involvement of this gene in CGHT pathogenesis.

To definitively implicate ABCA5 in CGHT pathogenesis, I tested the cosegregation of the splice variant with the disease phenotype and found indeed that both unaffected parents were heterozygous for this alteration. I characterized the effects of the substitution mutation on mRNA splicing by sequencing the transcripts and performing qRT-PCR on RNA from both affected and unaffected individuals, which revealed aberrant splicing activity that leads to nonsense-mediated decay and premature translational termination. I validated this finding at the protein level using immunoblotting in patient keratinocytes and fibroblasts, which revealed a marked decrease in both unmodified and modified (glycosylated) forms of the protein. I then investigated the cell type(s) synthesizing *ABCA5* in the human skin and hair follicle and used several

methods of detection, and thoroughly characterized its expression pattern in both healthy and affected tissues. Moreover, I investigated whether the ABCA5 expression pattern was conserved in mammals and using the same methods and approach in the analysis of mouse skin and hair follicles, observed strong expression in both epithelial and mesenchymal compartments in a pattern very similar to that of human hair follicles.

ABCA5 was previously localized with several other candidate genes with CGHT based on its genomic position (SUN *et al.* 2009; FANTAUZZO *et al.* 2012) but no intragenic mutations were previously reported in the chr17q24.2-24.3 candidate region or in any other genomic location. The identification of the *ABCA5* c.4320+1G>C splice site mutation and expression in human skin and hair follicles was novel at the time of this report. Concurrent with our studies in the autosomal recessive CGHT case, we analyzed a sporadic CGHT case of distinct ethnic origin, yet whose excessive hair overgrowth phenotype closely resembled that of the autosomal recessive CGHT patient. Considering the rare, sporadic nature of CGHT and the previous identification of CNVs in the chr17q24.2-24.3 region, we used a strategy very similar that described in Chapter II.2 in the sporadic XLH case to detect *de novo* structural variants with large pathogenic effects.

In sporadic CGHT, we sought to identify structural variants using cytogenetic methods including G-banding of chromosomes, FISH using BAC clones, sub-telomeric FISH, and performed CNV analysis using whole-genome mapping (DESTEFANO *et al.* 2014). These studies led to the identification of a

translocation between chromosomes 3 and 17 that was not present in the parents of the proband, and our CNV analysis revealed a cryptic 1.3 Mb deletion at one of the translocation breakpoints on chromosome 17, which overlapped with the CGHT candidate region and spanned *ABCA5.* In previous analyses studying the genetic basis of autosomal CGHT, the Christiano lab identified CNVs that created position effects on the downstream gene, *SOX9*, an important regulator of hair follicle growth in humans and mice (discussed in Chapter I.5). However, the lab had not previously investigated *ABCA5* as a candidate gene whose expression/function may be disrupted as a result of these CNVs.

In Chapter III.2, I postulated that *ABCA5* expression and/or function is compromised in the sporadic CGHT case, thus, I initially tested whether the second allele possessed an intragenic mutation in *ACBA5*. Using Sanger sequencing, I screened the exons and splice site junctions, as well as the promoter and regulatory sequences 2 Kb upstream of the gene, but did not identify mutations. I then tested the overall consequence of the 1.3 Mb deletion spanning *ABCA5* on the mRNA and protein levels of the gene, using the same approach and methods of detection I used in the previous case. Importantly, *ABCA5* expression was significantly reduced in patient hair follicles and the N-glycosylated form of the protein was not detectable, similar to what I observed in the autosomal recessive CGHT case, suggesting a loss-of-function mutation. The significance of this finding and potential mechanisms of *ABCA5* gene regulation are discussed in the following sections.

Upon the discovery of novel genes associated with a human disease phenotype, such as excessive hair overgrowth, the next step is to determine the functional consequences and mechanisms by which these genes act to control that trait. In Chapter III.2, I tested the consequence of *ABCA5* loss-of-function in CGHT at the subcellular level based on what was previously known about this transporter (discussed in the following section). Studies in mouse models have demonstrated a role for Abca5 in lysosome integrity in autophagosome formation and cholesterol efflux (KUBO *et al.* 2005; YE *et al.* 2010). Therefore, I determined whether these processes were defective in autosomal recessive CGHT. I performed an initial examination of lysosome function in collaboration with the Di Paolo laboratory and used a well-established assay to measure autophagosome formation and degradation in control and CGHT keratinocytes (described in Chapter III.2) (DESTEFANO *et al.* 2014). In parallel to these studies, we tested the distribution of free cholesterol in CGHT patient and control cells using a filipin stain (described in Chapter III.2) to determine whether the overall content and localization pattern was affected. Remarkably, the results from both of these studies revealed reduced lysosome function, accumulation of autophagosomes and their cargo, as well as increased endolysosomal cholesterol in CGHT, further demonstrating a loss-of-function mutation in ABCA5.

In Chapter III.2, I characterized the *ABCA5* loss-of-function mutation in CGHT and observed subcellular defects in lysosome function and cholesterol transport. However, the precise mechanism by which ABCA5 acts to control hair growth in health and disease remains unknown. Several *in vitro* and *in vivo*

functional studies for ABCA5 and other ABC transporters have been performed and may provide insightful hints as to what allocrites, or transported substrates, are effluxed from the cell in an ABCA5-dependent manner (QUAZI AND MOLDAY 2011). In the following sections, I discuss what is currently known about ABC transporters in health and disease, the molecules they transport, the evidence for ABCA5 in the transport of key lipids and hormones, and how these molecules influence human hair growth. Moreover, I propose ongoing and future experiments to more thoroughly characterize the allocrite(s) of this transporter and the mechanisms by which ABCA5 contributes to the excessive hair overgrowth phenotype observed in CGHT.

ABC transporters and their allocrites in human disease

ABC transporters are key for the movement of a variety of molecules and substrates across membrane lipid bilayers, and they possess two transmembrane domains, coupling function to ATP hydrolysis via the ATP-binding cassette motifs, Walker A, signature motif, and Walker B (QUAZI AND MOLDAY 2011). Functional studies of ABC transporters in the context of disease have demonstrated that the allocrite, or transported substrate, is the key determinant of the phenotype, as in some cases of human disease, the transporter itself is not expressed at the site of pathology (i.e. in the case of *ABCC6* in pseudoxanthoma elasticum) (RINGPFEIL *et al.* 2000).

Over the past decade, mutations in ABC transporters have been reported in several human syndromes (i.e. *ABCA1* in Tangier disease, *ABCA2* in

Alzheimer's disease, *ABCA4* in Stargadt disease, *ABCC6* in pseudoxanthoma elasticum, *ABCA12* in Harlequin ichthyosis, and *ABCC9* in Cantu syndrome) (QUAZI AND MOLDAY 2011; TARLING *et al.* 2013), where several of these pathologies manifest skin and hair phenotypes and are discussed in Chapter III.2. The large majority of ABC transporters are known/thought to mediate the ATP-dependent translocation of sterols, lipids, hormones and other related compounds, which are crucial for the integrity of a wide variety of biological processes. Interestingly, *ABCA5* is not well conserved among the ABCA family of transporters, possessing a maximal amino acid conservation of 42% among the chromosome 17q24 cluster (*ABCA5, ABCA6, ABCA8, ABCA9, ABCA10*), raising the question of whether these differences have biological impacts on protein localization and function.

As previously mentioned, *ABCA5* has never been associated with a human condition prior to the work presented in Chapter III.2, however, its upregulated expression has been observed in several tumors and is a reported biomarker for osteosarcoma, melanoma, undifferentiated colon and ovarian carcinomas (HEIMERL *et al.* 2007; OHTSUKI *et al.* 2007; SAINI *et al.* 2012). Notably, ABC transporters often confer resistance to drugs and chemotherapy/cancer-related substrates including phospholipids and cholesterol, therefore, it is likely that in the context of tumor biology, *ABCA5* upregulation may protect undifferentiated carcinoma cells from the intracellular accumulation of such toxic compounds.

Recently, a role for ABCA5 in cholesterol homeostasis was reported in neurons and in association with Alzheimer's disease. ABCA5 stimulated cholesterol efflux in neurons and reduced Aβ production *in vitro,* where the process of Aβ formation is thought to be regulated by membrane cholesterol (FU *et al.* 2015). *In vivo, ABCA5* expression was found to be elevated in the hippocampus of Alzheimer's disease patient brains, suggesting a novel role for this transporter in a neurological human pathology and further supporting the hypothesis that ABCA5 mediates the efflux of cholesterol.

Lipids, hormones and hair growth

In humans, several inherited hair disorders, including alopecia, autosomal recessive woolly hair (ARWH) syndrome, and localized autosomal recessive hypotrichosis 2 (LAH2) can be attributed to, at least in part by, defects in cholesterol or lipid trafficking (SHIMOMURA *et al.* 2008; SHIMOMURA *et al.* 2009a; SHIMOMURA *et al.* 2009b; SHIMOMURA AND CHRISTIANO 2010). Mutations in the genes, *P2RY5* and *LIPH* were initially identified from linkage and positional cloning studies in families with LAH2 and ARWH, yet at the time of their discovery, the precise mechanism by which these structural proteins acted to control human hair growth remained elusive. Thereafter, functional studies using biochemical assays, mouse models and *in vitro* cell culture systems demonstrated a role for lysophosphatidic acid (LPA) in keratinocyte differentiation and hair growth. *P2RY5*, or *LPAR6*, encodes a G protein-coupled receptor, which acts as a receptor for LPA, and *lipase H (LIPH)* encodes a phospholipase

enzyme that, through a cleavage event, produces LPA (SONODA *et al.* 2002; PASTERNACK *et al.* 2008). Mice null for *Liph* recapitulate pathological features of human WH syndrome (INOUE *et al.* 2011). Moreover, these functional studies elucidated the downstream effects of LPA signaling and convergent pathways, composed of the tumor necrosis factor converting enzyme (TACE), transforming growth factor (TGFα), and epidermal growth factor receptor (EGFR) pathways (INOUE *et al.* 2011).

The peroxisome proliferator-activated receptor gamma (PPARγ) plays a crucial role in lipid homeostasis within the skin and hair follicle (DI-POI *et al.* 2004), where it has multiple functions in epidermal differentiation, sebocyte differentiation and sebaceous gland formation, melanogenesis, dermal wound closure, as well as permeability barrier recovery (RAMOT *et al.* 2015). In mice, targeted ablation of PPARγ in the *K15* lineage leads to severe defects in hair and skin biology, producing a phenotype reminiscent of the human lichen planopilaris (LPP) condition characterized by dystrophic hair follicles and follicular plugging, massive fibrosis and sebaceous gland atrophy (KARNIK *et al.* 2009).

Beyond the context of lipids, hormone synthesis and signaling pathways have been described to influence human hair growth. Human patients with androgenetic alopecia (male-pattern baldness) progressively lose their hair on the front and top of their heads in response to increased androgen levels. Similarly, individuals with thyroid defects (both hypo- and hyperthyroidism) manifest skin and hair follicle abnormalities, including a thin epidermis and the development of alopecia. A recent study in mice examined the contribution of

thyroid hormone nuclear receptors TRα1 and TRα2 to hair follicle stem cell maintenance and found that mice deficient for these factors exhibit impaired epidermal proliferation, wound healing and hair growth, all of which were attributed to decreased stem cell activation in the hair follicle bulge (CONTRERAS-JURADO *et al.* 2015). Intriguingly, there was an overall increase in the number of label retaining cells (LRCs) and this accumulation perturbed Smad/β-catenin signaling during stem cell proliferation and mobilization, leading to hair loss.

There are many other examples of both humans and mice with dermatological defects resulting from increased/decreased hormone and lipid signaling. Intriguingly, all of the reported defects to date result in hair loss (alopecia), raising the question of what factors in these signaling pathways act as negative regulators in the hair follicle, such that their removal results in the production of more hairs.

Determining the function of ABCA5 in hair growth

Mouse *Abca5* expression has been reported in the brain, lung, heart and thyroid gland, where this transporter was found to localize to lysosomes and endolysosomal like structures at the subcellular level (KUBO *et al.* 2005). Moreover, this localization pattern was observed for the human and rat orthologues, whose cDNAs have been reported to be expressed in the brain, lung and testis (PETRY *et al.* 2003; PETRY *et al.* 2006).

Several years ago, *Abca5* deficient mice were generated and reported to possess defects in lysosome integrity, where the mice developed symptoms of

lysosomal storage like disease in the heart and thyroid glands (KUBO *et al.* 2005). During morphogenesis, there were no observable defects, but by postnatal day 30, symptoms of cardiomyopathy, exophthalmos and collapse of the thyroid gland were readily detectable, and the mice died shortly thereafter. Electron microscopy analysis of the affected tissues demonstrated an accumulation of autophasosome particles, highly suggestive of lysosome defect and consistent with the subcellular localization pattern of Abca5 (KUBO *et al.* 2005).

Interestingly, exophthalmos is a symptom often observed in Graves' disease in hyperthyroid patients, and *Abca5* deficient thyroid glands exhibit this defect (BAHN AND HEUFELDER 1993; BAHN 2002). Quantification of thyroid hormone levels in the mice revealed that *Abca5*$^{-/-}$ mice possess markedly reduced levels compared to their wild type littermates, contradicting the pathological nature of Graves' disease with increased thyroid levels and stimulated TSH receptor activity due to autoantibody production, and suggesting a distinct mechanism of action in *Abca5*$^{-/-}$ mice (BAHN AND HEUFELDER 1993; BAHN 2002). Importantly, Abca5 localizes to lysosomes, its deficiency is associated with the collapse of the thyroid gland, and lysosomes are known to be responsible for the secretion of thyroid hormones (RICHARDSON *et al.* 1996). Taken together, this evidence strongly supports a role for ABCA5 in the efflux of thyroid hormones, but further functional studies will be required to definitively implicate this transporter in the efflux of these and other substrates.

It currently remains unknown whether *Abca5* deficient mice possess an excessive hair overgrowth phenotype because this was not directly tested.

Further detailed histological analyses in these mice along with additional experiments (discussed in Chapter III.2) will elucidate the contribution of Abca5 to regulating hair growth. Upon a cursory examination of the *Abca5*$^{-/-}$ mice, it is clear that the mutant adult mice possess a full coat of hair, which is in contrast to the phenotype observed in human patients with hypo- or hyperthyroidism (discussed in the following section) (KUBO *et al.* 2005). This observation, in conjunction with the genetic association of human *ABCA5* with CGHT, highlights the possibility that *Abca5* loss-of-function in mice may promote hair growth rather than hair loss.

In a separate study, the function of Abca5 was interrogated in the context of lipid signaling and cholesterol transport (YE *et al.* 2010), as several other members of the *Abca* family were previously implicated in the transport of cholesterol and other substrates and previous reports have demonstrated strong expression of *Abca5* in the Leydig cells of the testis (a site of sterol hormone synthesis) (PETRY *et al.* 2006). Abca5 was found to function in resident macrophage cells of the liver, where its loss-of-function perturbed cholesterol efflux and led to the development of atherosclerosis following bone marrow transplantation of *Abca5*$^{-/-}$ macrophages into female irradiated LDLr$^{-/-}$ mice (YE *et al.* 2010).

Collectively, the evidence based on Abca5 localization at the subcellular and tissue levels, as well as phenotypes observed in *Abca5*$^{-/-}$ mice strongly support a role for Abca5 in the endolysosomal system involved in the efflux of key sterols, hormones and/or lipids. We postulate that ABCA5 loss-of-function in

CGHT perturbs cholesterol and/or thyroid hormone transport during morphogenesis, disrupting key signaling pathways required for normal hair follicle patterning, morphogenesis and growth.

Ongoing and future experiments

Several *in vitro* and *in vivo* studies can be performed to test the role of *ABCA5* in CGHT disease pathogenesis and identify the allocrite(s) whose perturbed transport mediates phenotypic effects. Using the *Abca5*$^{-/-}$ mouse model, the morphology, distribution and length of hairs can be measured using histological and morphometric analyses. Although a hair phenotype was not initially reported in these mice, hair phenotypes in mutant mice deficient for several factors can be subtle in the absence of detailed histological analyses. It is important to test all hair types (pelage hairs, vibrissae, tail hair), as they are distinct from one another in the context of morphology and genetic control (DUVERGER AND MORASSO 2009). Moreover, it is known that mouse models for human hair conditions do not always possess the same (or expected) phenotype. For example, *FGF5* mutations cause excessive hair overgrowth of mouse pelage hairs but only affect the eyelashes in humans. Likewise, the *Koala* mouse model has excessive hair overgrowth on the ears and muzzle, in contrast to the human Ambras syndrome patients. (FANTAUZZO *et al.* 2008; HIGGINS *et al.* 2014). Inducible overexpression of *Abca5* during morphogenesis can be performed followed by phenotypic analysis to determine whether these mice exhibit features associated with

mutations in cholesterol and/or thyroid hormone transport-related proteins in mice, leading to hair loss.

To determine the contribution of Abca5 to hair growth in one particular lineage (i.e. epithelial), conditional deletion of Abca5 can be performed in mice by crossing a floxed allele with a tissue-specific driver (i.e. *K14* for epithelial cells of the skin and hair follicle). Testing the contribution of sterols in the skin during or following morphogenesis can be accomplished by topical administration of cyclodextrin, followed by histological analyses. Moreover, blocking ABCA transporter activity using glyburide treatment or other pharmacological inhibitors will elucidate whether the transporter positively or negatively regulates hair growth. Several biochemical assays and methods that have been previously used to identify the allocrites of other transporters include fluorescence anisotrophy using purified recombinant protein (measure binding affinity to transporter extracellular domain), *in vitro* transport assays using semi-permeabilized cells, screening peptide libraries, and trapping transported peptides in the ER via glycosylation and ATP-independent binding assays (ANDROLEWICZ *et al.* 1993; HENDERSON *et al.* 1993; SUN *et al.* 1999; BISWAS-FISS *et al.* 2012). Given the evidence for ABCA5 in transporting cholesterol and thyroid hormones, these substrates can be directly tested and validated in *Abca5*$^{-/-}$ cells combined with competition assays. Additional functional studies using human patient materials can provide insight into the role of ABCA5 in thyroid hormone and cholesterol transport by measuring the overall levels of

these molecules and their subcellular localization patterns in ABCA5 loss-of-function cells those rescued with the full-length coding sequence of ABCA5.

A possibility to be considered in the pathogenesis of autosomal recessive CGHT is the contribution of other rare variants and their genetic interactions with *ABCA5* to the overall disease phenotype. A plausible candidate that fits with this hypothesis is the *DGKZ* gene that encodes diacylglycerol (DAG) kinase zeta and whose protein phosphorylates phosphatidic acid, plays important roles in lipid signaling, and is implicated in neurological diseases (PACHECO AND JOPE 1996; VAN BLITTERSWIJK AND HOUSSA 2000). The missense mutation (c.1678C>T(p.P560S)) cosegregated with the CGHT disease phenotype in a homozygous recessive manner and in my initial expression analysis, I observed detectable levels of *DGKZ* transcripts in human skin and hair follicles. Of interest, the autosomal recessive CGHT patient also has epilepsy, and mice deficient for another DAG kinase gene (*Dgke*) exhibit features associated with epilepsy (RODRIGUEZ DE TURCO *et al.* 2001). This association, in conjunction with the recently reported role of ABCA5 function in the brain (FU *et al.* 2015), it is likely that the neurological phenotypes associated with CGHT in this patient are attributed to both *ABCA5* and *DGKZ* loss-of-function mutations. This hypothesis can be tested in these mouse models, generating double mutants and compound heterozygotes, as well as by performing *in vitro* rescue experiments.

In the context of gene regulation, we found in the sporadic CGHT case that ABCA5 levels were significantly reduced to those observed in the autosomal recessive CGHT case, suggesting the loss-of-function by haploinsufficiency.

Plausible mechanisms to explain this phenomenon are discussed in Chapter III.2 and include monoallelic expression of this gene/locus, imprinting, *trans* regulatory interactions between homologous chromosomes, and dosage sensitivity. Testing monoallelic expression can be accomplished by overlaying SNP data with transcriptome sequencing (SKELLY *et al.* 2011), and in the context of imprinting, I examined the list of genes identified as being maternally imprinted from a screen performed in *Dnmt3L* mutant mice and did not identify *ABCA5* as an imprinted gene (SCHULZ *et al.* 2010). *Trans* regulatory interactions can be identified using the Chromosome Conformation Capture (3C) technology with high-throughput sequencing, and differences in chromatin states between alleles can be determined using the Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE) assay (SIMON *et al.* 2012), combining this ChIP-based approach with SNP genotyping to distinguish alleles.

Lastly, dosage sensitivity of *ABCA5* could be determined through the use of mouse models comparing the phenotypes between *Abca5+/-* to *Abca5+/+* and *Abca5-/-* mice, and also by deleting the locus from one allele and assessing the downstream consequences. In sporadic CGHT, an analysis of lysosome function and cholesterol transport could provide insight as to whether deletion of one copy of *ABCA5* is sufficient to produce a phenotype. However, unaffected related family members in the autosomal recessive CGHT case harbor one copy of the loss-of-function allele and do not have hypertrichosis, suggesting distinct mechanisms contribute to decreased *ABCA5* expression in sporadic CGHT.

Chapter IV

Functional genomics at the *ULBP6* locus identifies a role for CTCF variants in alopecia areata

(Manuscripts #3-5)

Chapter IV.1 Preface

In the study of non-Mendelian, complex disorders, the methods and approaches used to identify pathogenic variants are distinct from those employed for Mendelian or monogenic diseases. Several reasons why common disorders are not amenable to conventional approaches include the lack of a clear inheritance pattern within families displaying the trait, the involvement of many variants that are more common within the general population and produce small effects, and the contribution of several environmental factors to the penetrance and severity of the disease. Therefore, the informative value of parametric linkage studies and genome-wide scans in family cohorts was limited because of the difficulty in identifying multiplex families in which the disease segregated. As a result, the most powerful approach to study these common conditions was to search for loci that are significantly associated across a population of unrelated patients (rather than shared within a family).

Genome-Wide Association Studies (GWAS) were developed to identify common variants using a genotyping-based approach in case-control cohorts and identified new susceptibility loci. The success of a GWAS depends largely on the nature of the disease phenotype, variation, as well as sensitivity and specificity in the diagnosis. While the GWAS approach has been widely used for hundreds of complex diseases and traits, it operates under the assumption that these diseases are caused by many common variants with small effects and are present at a frequency of 5% or greater in the general population. As such, the proportion of heritability of a disease that has been accounted for by identified

variants identified in GWAS has been relatively small, highlighting the limitations of this method in detecting the rarer variants with larger effects, structural variants, and genetic interactions we now know contribute significantly to the "missing heritability." Interestingly, the majority of susceptibility variants identified from GWAS reside within non-coding, intergenic DNA, suggesting that these variants predominantly exert pathogenic effects by disrupting regulatory regions and transcription factor binding sites.

To circumvent the limitations of a GWAS, a robust and comprehensive approach must be used to detect the pathogenic variants in GWAS studies, and somehow incorporate this information into an understanding of the genetic architecture of the disease. A major challenge is how do we know what loci are biologically interesting and worth following up on? For genomic regions significantly associated with a trait, it is optimal to confirm these loci in an independent replication study, and subsequently, in a meta-GWAS analysis. Moreover, it is key to have some functional knowledge of the genes encompassed in the LD blocks and whether there is biological relevance to the disease phenotype. This information can be gathered from a variety of sources including prior knowledge of the gene(s) associated with the trait (family-based linkage studies, CNV analysis), mutations identified in similar Mendelian disorders (if applicable), pathological features of the disease that correlate with the dysregulation of certain genes and signaling pathways, eQTL analysis, and studies in mouse models for the disease. Based on this information, susceptibility

loci can then be prioritized for targeted deep resequencing to capture all the coding and non-coding variants in the candidate GWAS regions.

The completion of the Human Genome Project and development of next-generation sequencing methods have been instrumental in the recent implementation of the functional genomics approach to studying complex human disorders. Moreover, the ENCODE and Roadmap Epigenomics projects have greatly facilitated the identification of novel DNA regulatory elements, such that functional genomics studies can be used to test the consequences of disease-associated variants in these regions (described in Chapter IV.4). Importantly, this approach has successfully been used for several complex genetic disorders to elucidate the mechanisms by which non-coding DNA can influence the regulation of gene expression levels (HUTCHINSON *et al.* 2014) (VERLAAN *et al.* 2009).

In the work described in Chapter IV, we functionally decoded the genetic architecture of the *ULBP3/6* locus in alopecia areata using targeted resequencing, identification of pathogenic variants, and determined their functional consequences. The Christiano lab previously performed the first GWAS in AA and identified eight regions of the genome that cluster into discrete LD blocks. As is observed for most autoimmune disorders, the *HLA* region was identified as the most significantly associated locus, but outside this region is a novel locus on chromosome 6q25.1 encoding the *ULBP3/6* genes that is strongly associated with the disease. Importantly, the *ULBP* genes encode ligands for the cytotoxic T cell receptor, NKG2D that has previously been implicated in AA

pathogenesis, and these ligands possess strong upregulation in human AA lesional hair follicles.

In Chapter IV.2, we replicated the findings from the initial GWAS study and identified new loci associated with the disease. We reported a replication study and meta-GWAS analysis conducted for AA that validated the previous associations, and moreover, identified two novel loci associated with AA. This work was performed in collaboration with Dr. Regina Betz and her research group at Bonn University (Germany) who provided the independent population cohorts (controls and patients) of European ancestry (author contributions described in Chapter IV.2). In the work described in Chapter IV.2, I characterized a new locus identified on chromosome 2q13, encoding the *BCL2L11 (BIM)* gene, that functions in apoptosis and the immune system. I first determined its expression levels in healthy human immune cells. To determine whether this gene was expressed in the hair follicle, I characterized the expression pattern of BIM in both human and mouse hair follicles using immunofluorescence staining, and performed immunohistochemical stains on the C3H/HeJ mouse model to detect BIM localization in AA lesional hair follicles. Based on my findings and the known role of this pro-apoptotic protein in the immune system, I postulated that dysregulation of BIM in affected individuals may contribute to the premature entry of AA hair follicles in dystrophic catagen that is characterized by massive apoptosis during active disease.

In parallel with these studies, we also defined the immunological features underlying both human and mouse AA, based on the GWAS and meta-GWAS

findings implicating genes of both the adaptive and innate immune system. In Chapter IV.3, we collaborated with Dr. Raphael Clynes and members of his laboratory to investigate the role and contribution of the NKG2D-ULBP signaling pathway to AA disease pathogenesis. Initially, we used the mouse model to characterize the immunophenotype of the immune infiltrate in lesional hair follicles using immunofluorescence staining, FACS and cytotoxicity assays. Importantly, we utilized the grafting method in the C3H/HeJ mouse model to induce autoimmune hair loss, which allowed us to study disease prevention as well as reversal. To determine the functional requirement of CD8+NKG2D+ T cells, we performed subcutaneous injections of lymph node cells in unaffected and affected AA mice and determined that these cells were necessary and sufficient to induce disease pathogenesis. These experiments were performed by Drs. Zhenpeng Dai, Luzhou Xing and Ali Jabbari.

In Chapter IV.3, I characterized the molecular profiles of human and mouse AA using microarray and RNA-seq technologies for transcriptional profiling on affected vs. unaffected skin. In this work, I analyzed the expression data from the mouse model and validated the dysregulation of key immune genes in these pathways using RT-PCR and immunofluorescence staining. We next used bioinformatics tools and statistical analyses to distinguish three primary gene expression signatures that included the interferon response genes, markers of cytotoxic T lymphocytes, and gamma chain cytokines, where Drs. Lynn Petukhova and Jane Cerise performed this work. Importantly, this transcriptional information correlated with genetic data from the GWAS, and

allowed us to define the signaling cascade involving the NKG2D-ULBP signaling pathway to identify the downstream genes and pathways as targets for therapeutic interventions. Using small molecule inhibitors (the FDA-approved drugs Tofacitinib and Ruxolitinib), we blocked the JAK/STAT signaling pathway in the mouse model to both prevent as well as reverse disease pathogenesis, and these experiments were conducted by Drs. Zhenpeng Dai, Luzhou Xing and Ali Jabbari.

Considering the findings obtained from our functional studies in the mouse model, we next tested whether targeting the JAK/STAT pathway with these small molecule inhibitors could treat alopecia areata in human patients. Thus, we conducted a pilot clinical trial using Ruxolitinib and successfully treated three AA patients with Drs. Ali Jabbari and Julian Mackay-Wiggan obtained the patient samples for this study and performed the analyses. Importantly, our integrative approach to define the molecular pathology of AA was greatly informed from the initial GWAS findings, which implicated key genes and signaling pathways in disease pathogenesis. Furthermore, our expression analyses and genome-wide transcriptional profiling, together with our functional studies in the mouse model led to the identification of downstream targets as therapeutic interventions for both disease prevention and reversal in mouse and human AA.

In Chapter IV.4, I investigated the central question of what genes and mechanisms control spatiotemporal expression in the context of hair follicle immune privilege. Informed from the GWAS results and functional studies performed in mouse and human AA, I focused my studies on the *ULBP3/6* locus

on chr.6q25 and sought to characterize the pathogenic variants and underlying mechanisms by which they contribute to dysregulation of the *ULBP3/6* genes in AA lesional hair follicles. In Chapter IV.4, we performed targeted deep resequencing of this locus, and using functional genomics, identified common and rare variants that are enriched on the GWAS risk haplotypes. We annotated the non-coding variants using ENCODE and RegulomeDB to determine whether they interrupted TFBS and/or regulatory regions. These studies and analyses were conducted by Drs. Lynn Petukhova and Esther Drill. Given the emerging view that the genetic architecture underlying common disease is composed largely of non-coding variants, and that rare variants driving genetic associations produce larger effects than the common variants, I focused my subsequent studies on the rare, non-coding variants identified at the *ULBP3/6* locus by targeted deep sequencing.

In the work presented in Chapter IV.4, I aimed to characterize three rare and novel non-coding variants we identified that reside within CTCF binding sites downstream of *ULBP6*. I postulated that non-coding variants contribute to dysregulated transcription factor binding, regulatory activity, and/or long-range interactions at the *ULBP3/6* locus. To test this hypothesis, I examined endogenous CTCF binding at the *ULBP6* region *in vivo* using ChIP-PCR on healthy human scalp dermal fibroblasts, and thereafter characterized the regulatory activity of this region using luciferase reporter assays. I next tested the effects of the rare, non-coding variants on reporter activity, and investigated whether these variants can disrupt CTCF binding by overexpressing a tagged

form of CTCF and measuring reporter activity *in vitro*. To determine whether these variants created novel transcription factor binding sites, I performed an *in silico* analysis of the *ULBP6* control and AA sequences.

In the context of alopecia areata and the loss of immune privilege in the hair follicle, I postulated that the *ULBP3/6* locus may be spatiotemporally regulated and engages in CTCF-mediated long-range interactions. Consistent with my hypothesis, we found that there are several clustered sites of CTCF binding enrichment across the locus and that the non-coding variants overlap with interacting sequences from ENCODE CTCF-mediated Chromatin Interaction with Paired-End Tags (ChIA-PET) experiments. Future studies will test whether this candidate region is capable of engaging in long-range interactions *in vitro,* and identify CTCF-mediated long-range interactions at this locus that may be lost in human AA lesional hair follicles.

Collectively, our strategy and approach to define the complex genetic architecture of alopecia areata was successful in identifying key genes and regulatory pathways that can be targeted to treat disease. Moreover, using functional genomics to study the *ULBP3/6* locus allowed us to interrogate the effects of disease-associated rare, non-coding variants and identify the mechanisms by which they exert their pathogenic effects.

Manuscript Contributions

Chapter IV.2. Genome-wide meta-analysis in alopecia areata resolved HLA associations and reveals two new susceptibility loci (Manuscript #3).

In this study, I characterized the expression of BIM, GARP, and LNK in human immune cell populations, including T cells, B cells, monocytes, NK cells, and PBMCs, as well as scalp hair follicles using RT-PCR. I then analyzed the expression pattern of BIM in human hair follicles as well as mouse hair follicles in anagen, catagen and telogen using immunofluorescence staining and confocal microscopy. Furthermore, I performed immunofluorescence staining for BIM on C3H/HeJ AA affected and unaffected hair follicles.

Chapter IV.3. Genome-wide meta-analysis in alopecia areata resolved HLA associations and reveals two new susceptibility loci (Manuscript #4).

In this work, I performed microarray expression profiling on total RNA isolated from C3H/HeJ affected and unaffected skin and analyzed the gene expression data. I validated dysregulated expression of key transcripts in immune signaling pathways using qRT-PCR as well as immunofluorescence staining on AA affected and unaffected skin.

Chapter IV.4. Functional genomics at the *ULBP6* locus identifies a role for CTCF variants in alopecia areata (Manuscript #5).

In this study, I analyzed the function of the three rare variants we identified from targeted deep resequencing that reside within CTCF binding sites. I performed Sanger sequencing for these variants on patient DNA samples and cloned both control and AA sequences (250 bp region) into a reporter vector to characterize the regulatory activity of this region. I determined endogenous CTCF binding at the *ULBP6* locus using ChIP-PCR on human dermal fibroblasts, and in these same cells, overexpressed a tagged form of CTCF in the presence of the AA rare variants to test changes in binding affinity. Moreover, I performed *in silico* analysis of the candidate region to identify novel transcription factor binding sites.

Chapter IV.2

Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci

(Manuscript #3)

nature COMMUNICATIONS

ARTICLE

Received 20 Jun 2014 | Accepted 25 Nov 2014 | Published 22 Jan 2015

DOI: 10.1038/ncomms6966

Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci

Regina C. Betz[1,*], Lynn Petukhova[2,3,*], Stephan Ripke[4,5,*], Hailiang Huang[4,5], Androniki Menelaou[6], Silke Redler[1], Tim Becker[7,8], Stefanie Heilmann[1,9], Tarek Yamany[2], Madeliene Duvic[10], Maria Hordinsky[11], David Norris[12], Vera H. Price[13], Julian Mackay-Wiggan[2], Annemieke de Jong[2], Gina M. DeStefano[14], Susanne Moebus[15], Markus Böhm[16], Ulrike Blume-Peytavi[17], Hans Wolff[18], Gerhard Lutz[19], Roland Kruse[20], Li Bian[2], Christopher I. Amos[21], Annette Lee[22], Peter K. Gregersen[22], Bettina Blaumeiser[23], David Altshuler[4,5], Raphael Clynes[2,24], Paul I.W. de Bakker[6,25], Markus M. Nöthen[1,9], Mark J. Daly[4,5] & Angela M. Christiano[2,14]

Alopecia areata (AA) is a prevalent autoimmune disease with 10 known susceptibility loci. Here we perform the first meta-analysis of research on AA by combining data from two genome-wide association studies (GWAS), and replication with supplemented ImmunoChip data for a total of 3,253 cases and 7,543 controls. The strongest region of association is the major histocompatibility complex, where we fine-map four independent effects, all implicating human leukocyte antigen-DR as a key aetiologic driver. Outside the major histocompatibility complex, we identify two novel loci that exceed the threshold of statistical significance, containing ACOXL/BCL2L11(BIM) (2q13); GARP (LRRC32) (11q13.5), as well as a third nominally significant region SH2B3(LNK)/ATXN2 (12q24.12). Candidate susceptibility gene expression analysis in these regions demonstrates expression in relevant immune cells and the hair follicle. We integrate our results with data from seven other autoimmune diseases and provide insight into the alignment of AA within these disorders. Our findings uncover new molecular pathways disrupted in AA, including autophagy/apoptosis, transforming growth factor beta/Tregs and JAK kinase signalling, and support the causal role of aberrant immune processes in AA.

[1] Institute of Human Genetics, University of Bonn, Bonn D-53127, Germany. [2] Department of Dermatology, Columbia University, New York, New York 10032, USA. [3] Department of Epidemiology, Columbia University, New York, New York 10032, USA. [4] Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. [5] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02114, USA. [6] Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. [7] German Center for Neurodegenerative Diseases, Bonn D- 53175, Germany. [8] Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn D-53127, Germany. [9] Department of Genomics, Life and Brain Center, University Bonn, Bonn D-53127, Germany. [10] Department of Dermatology, MD Anderson Cancer Center, Houston, Texas 77030, USA. [11] Department of Dermatology, University of Minnesota, Minneapolis, Minnesota 55455, USA. [12] Department of Dermatology, University of Colorado, Denver, Colorado 80204, USA. [13] Department of Dermatology, University of California, San Francisco, San Francisco, California 94110, USA. [14] Department of Genetics and Development, Columbia University, New York, New York 10032, USA. [15] Institute of Medical Informatics, Biometry, and Epidemiology, University Duisburg-Essen, Essen D-45122, Germany. [16] Department of Dermatology, University of Münster, Münster D-48149, Germany. [17] Department of Dermatology and Allergy, Clinical Research Center for Hair and Skin Science, Charité-Universitätsmedizin Berlin, Berlin D-10117, Germany. [18] Department of Dermatology, University of Munich, Munich D-80337, Germany. [19] Dermatological Practice, Hair and Nail, Wesseling D-50389, Germany. [20] Dermatological Practice, Paderborn D-33098, Germany. [21] Community and Family Medicine and Genetics, Dartmouth College, Hanover, New Hampshire 03755, USA. [22] The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA. [23] Department of Medical Genetics, University of Antwerp, Antwerp BE-2650, Belgium. [24] Department of Medicine, Columbia University, New York, New York 10032, USA. [25] Department of Epidemiology, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.M.C. (email: amc65@columbia.edu).

Alopecia areata (AA) is one of the most prevalent autoimmune diseases, with a lifetime risk of 1.7% (ref. 1), and is the most common cause of hair loss in children. In AA, aberrant immune destruction is targeted to the hair follicle, resulting in non-scarring hair loss that typically begins as patches, which can increase in size and coalesce and may progress to cover the entire scalp (alopecia totalis), and body as well (alopecia universalis). Disease prognosis is unpredictable and highly variable. Its aetiologic basis has remained largely undefined, creating barriers to the development of effective therapeutic strategies and resulting in an enormous unmet medical need[2,3].

Our first GWAS in AA identified associations in eight regions of the genome, which were subsequently confirmed in independent candidate gene studies[4–7]. Associated loci outside the human leukocyte antigen (HLA) highlight particular immune response pathways and also implicate genes expressed in the hair follicle. For example, several regions contain genes with Treg functions, including *IL2RA*, *IL2/IL21*, *CTLA4* and *Eos*, whereas *ULBP3/ULBP6* implicate NKG2D-mediated cytotoxic T cells. Within the hair follicle, expression of *STX17* suggests a role for end-organ autophagy, while *PRDX5* implicates oxidative stress. A combined analysis of this GWAS and a subsequent replication study led to the identification of *IL13/IL4* and *KIAA0350/CLEC16A* as new gene loci[5].

Here we perform a meta-analysis to expand our sample size, and identify two new loci that exceed our threshold for genome-wide significance and a third locus that is nominally significant. We identify transcripts and/or protein for candidate genes at all three loci in disease-relevant tissues. We perform imputation and fine-mapping of the HLA, identifying four independent associations that implicate HLA-DRβ1. Finally, a cross phenotype meta-analysis (CPMA) of our data with published results from seven other autoimmune diseases identify molecular pathways shared by AA and one or more other disorders.

Results

In this study, we have increased our cohort size and performed a combined analysis of two GWAS using Illumina Human660W- and Omni1-Quad BeadChips, analysing a total of 2,489 cases and 5,287 controls ascertained in the US and Central Europe (Supplementary Table 1). Association analyses are performed with logistic regression. In a meta-analysis of these data, nine of the previously implicated regions exceeded statistical significance ($P<5\times10^{-8}$), with *STX17* achieving nominal significance (rs10124366; $P=1.09\times10^{-5}$) (Fig. 1; Supplementary Data 1).

First, to resolve the major histocompatibility complex (MHC) association signal ($P=4.91\times10^{-58}$ for the best single-nucleotide polymorphism (SNP), rs9275516), we used a published imputation and analysis protocol to perform fine-mapping (Supplementary Data 2)[8]. Conditional analysis revealed four independent variants located at the classical *HLA-DRA* and *HLA-DRB1* genes. The most significant variant was amino-acid position 37 in HLA-DRβ1 (omnibus P value $=4.99\times10^{-73}$). Of the five possible amino acids at this position, Leu (odds ratio (OR) = 1.56), Tyr (OR = 1.54) and Phe (OR = 1.19) conferred a higher risk of AA, whereas the other residues conferred lower risk (OR for Asn = 0.42; OR for Ser = 0.74). Adjusting for the effects of HLA-DRβ1 amino-acid position 37, we found an independent association due to an intronic SNP of *HLA-DRA*, rs9268657 (OR = 0.63, P value $=4.48\text{E}^{-41}$). Functional annotation of this SNP and its close proxies ($R^2>0.9$) reveal that they influence expression levels of *HLA-DRB1* (ref. 9). Adjusting for both HLA-DRβ1 amino-acid position 37 and rs9268657, we identified another independent association for the classical allele HLA-DRB1*04:01 (OR = 1.64, P value $=1.76\times10^{-12}$), confirming previous associations in candidate gene studies[10,11]. Finally, adjusting for HLA-DRβ1 amino-acid position 37, rs9268657 and HLA-DRB1*04:01, amino-acid position 13 in HLA-DRβ1 was also significant (P value $=4.57\times10^{-16}$). Of the six possible alleles at this position, three increased risk (Tyr, OR = 1.41; Ser, OR = 1.35; Phe, OR = 1.09), two are protective (His, OR = 0.57; Gly, OR = 0.50) and one demonstrated no effect (Arg, OR = 0.98) (Supplementary Table 2). Further rounds of conditional analyses yielded no additional significant results ($P>2.11\times10^{-6}$).

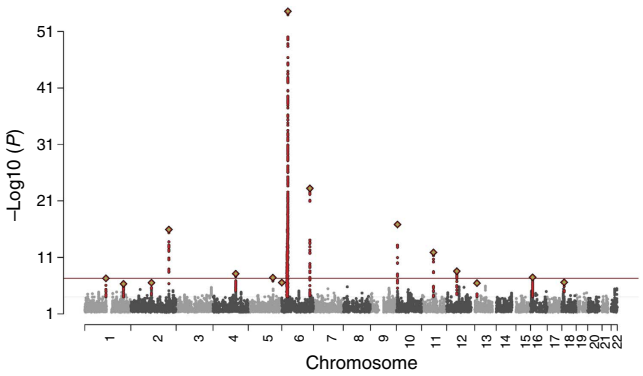


Figure 1 | Manhattan plot for genome-wide tests of association in meta-analysis. To conduct a meta-analysis across two GWAS, genotypes were imputed for each data set yielding 1.2 million SNPs. Standard association analysis with logistic regression including PC covariates was performed within each cohort and results were combined with s.e. weighted meta-analysis.

Collectively, these four independent associations in the MHC implicate HLA-DR as the primary risk factor in AA, presumably through antigen presentation, similar to other immune-mediated diseases[12]. For example, HLA-DRβ1 amino-acid positions 13 and 37 contribute to P4 and P9 peptide-binding pockets, respectively; the disease associations of HLA-DRB1*04:01 are thought to be driven by the shared epitope (amino-acid residues 70–74)[12], which also occur within the peptide-binding cleft (Fig. 2). Polymorphic residues within peptide-binding pockets of HLA-DR influence binding affinities of peptides, and thus may shape the repertoire of autoantigens capable of triggering or perpetuating disease.

Next, we performed replication of SNPs in functionally relevant loci that achieved suggestive evidence for association in the meta-analysis ($5\times10^{-8}>P>1\times10^{-5}$), utilizing an independent cohort of Central European ancestry. The Immunochip was used to genotype 318 cases and 1,688 controls and a Sequenom assay was used to genotype 85 SNPs not included on the Immunochip in a cohort of 764 cases and 568 controls. This analysis identified statistically significant associations in two novel genomic regions: chromosome 2q13 containing *ACOXL* and *BCL2L11* (rs3789129, $P=1.51\times10^{-8}$, $OR_A=1.3$), and chromosome 11q13.5 containing *C11orf30* and *LRCC32* (rs2155219, $P=1.25\times10^{-8}$, $OR_T=1.2$) (Table 1; Supplementary Data 3).

The association signal at chromosome 2q13 is located within an intron of *ACOXL*, and the region of association extends to include *BCL2L11* (Fig. 3). This region has been implicated in GWAS for two other autoimmune diseases: immunoglobulin-A nephropathy and primary sclerosing cholangitis[13]. *ACOXL* belongs to the acyl-coenzyme A oxidase gene family. While other family members play well-studied roles in peroxisomal beta-oxidation, very little is known about the function of this

gene. BCL-2-like 11, also known as BIM, is a member of the BCL-2 protein family and contains a Bcl-2 homology domain 3 (BH3) that interacts with other members to act as a pro-apoptotic factor. BIM has been widely studied in the context of apoptosis of immune cells, and has also been implicated in apoptosis of melanocytes[14]. More recently, data have emerged that implicate BIM in the destruction of some end-organs in autoimmune diseases, such as pancreatic beta cells in type 1 diabetes[15]. Furthermore, BIM has also been shown to regulate autophagy[16], adding to growing evidence for the importance of this process in AA as recently shown by functional studies of STX17 (refs 17,18).

BIM is widely expressed in multiple cell types, including immune-related, epithelial and hair follicle cells[19]. We first performed reverse transcription (RT)–PCR analysis of *BIM* expression in immune cells and scalp hair follicles (SHFs). We detected *BIM* transcript in whole-peripheral blood mononuclear cells (PBMCs), including T, natural killer and B cells, and monocytes (MCs) (Fig. 4a). In plucked human SHFs, we observed expression of an alternative splice isoform, *BIM-S*, the most pro-apoptotic variant (Fig. 4a)[20,21]. To characterize protein localization within hair follicles, we performed immunofluorescence staining on both human and mouse hair follicles. In the human hair follicle, BIM is highly expressed in the bulb of the hair follicle, especially within the matrix cells, in a pattern strikingly restricted to early catagen, but not anagen or telogen (Fig. 4b). The hair bulb matrix is the principal location of differentiated, pigment-producing melanocytes, which are postulated to undergo apoptosis during catagen[22]. In mouse hair follicles, BIM localized to the apoptosing strand and lower portion of the catagen hair follicle, consistent with its expression in human hairs. Remarkably, BIM expression was restricted to catagen hair follicles, where immunofluorescence staining performed on anagen or telogen hairs produced no signal (Fig. 4c, upper panel). Finally, to investigate a possible role of BIM in AA pathogenesis, we used a mouse model that recapitulates genetic and molecular profiles of human AA[23]. C3H/HeJ spontaneously develops AA, and we performed immunofluorescence staining on affected and unaffected skin. We observed a striking and widespread increase in BIM expression throughout affected skin and hair follicles, which was not restricted to catagen hair follicles and appeared further upregulated in lesional skin (Fig. 4c, lower panel). The regression stage of a normal cycling hair follicle, catagen, is characterized by apoptosis and cell death. We postulate that dysregulation of BIM in hair-matrix keratinocytes and/or hair follicle melanocytes contributes to the early entry into dystrophic catagen in AA hair follicles in active disease.

The second novel region that exceeded statistical significance at chromosome 11q13.5 (rs2155219, $P = 1.25 \times 10^{-8}$) contains three gene transcripts, including C11orf30, GUCY2E and GARP (LRRC32). GWAS have implicated this region in several autoimmune and inflammatory diseases (www.genome.gov/gwastudies). GARP is expressed on activated Tregs and binds and augments transforming growth factor beta bioavailability[23–26], functioning to induce FOXP3 expression and Treg differentiation. Expression of GARP on activated human Treg cells correlates with

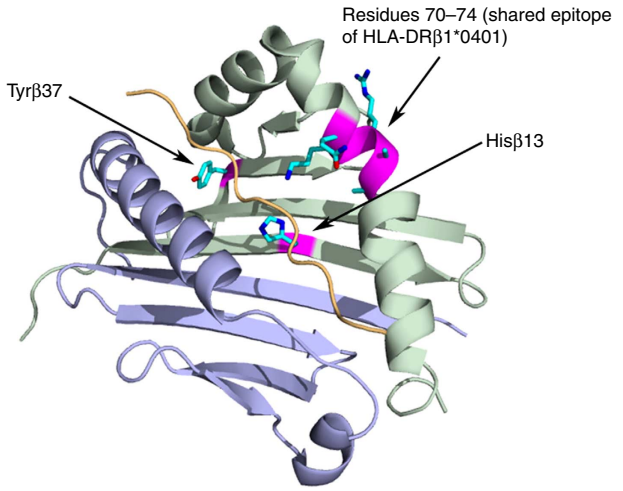


Figure 2 | Positions of amino-acid residues demonstrating independent association with alopecia areata. Structure of disease-associated HLA-DRB1*04:01 allele and polymorphic residues involved in susceptibility to AA. The peptide-binding cleft of an HLA-DR molecule is shown in cartoon representation with the α-chain coloured in blue and β-chain in green. The MHC-bound peptide is shown in cartoon representation and coloured yellow. Residues His13β and Tyr37β corresponding to amino-acid positions with independent disease association are shown in a stick model along with the shared epitope (residues 70–74). Crystal structure of HLA-DR4 bound to melanocytes lineage-specific antigen gp120 was used (PDB 4IS6).

Table 1 | Candidate genes in AA GWAS regions.

Locus	Genes of interest	SNP	Chr	BP	A1A2	*P*	OR
6p21.32	HLA-DQB1	rs9275524	6	32,783,087	TC	**1.8E−60**	0.52
10p15.1	IL15RA,IL2RA	rs3118470	10	6,141,719	TC	**7.7E−21**	0.71
2q33.2	CD28,CTLA4,ICOS	rs231775	2	204,440,959	AG	**2.2E−20**	0.72
6q25.1	RAET1L,ULBP3	rs12183587	6	150,396,301	TG	**5.9E−24**	1.48
11q13	PRDX5	rs574087	11	63,859,524	AG	**8.7E−14**	1.32
12q13	IKZF4 (Eos),ERBB3	rs2292239	12	54,768,447	TG	**4.4E−09**	1.25
4q27	IL21,IL2	rs7682481	4	123,743,476	CG	**4.8E−09**	1.23
5q31.1	IL13,1L4	rs848	5	132,024,399	AC	**4.8E−09**	1.27
2q13	ACOXL,BCL2L11(BIM)	rs3789129	2	111,414,511	AC	**1.5E−08**	1.31
11q13.5	GARP(LRRC32)	rs2155219	11	75,976,842	TG	**4.1E−08**	1.21
1p13.2	PTPN22	rs2476601	1	114,179,091	AG	8.9E−08	1.34
12q24.12	SH2B3(LNK),ATXN2	rs653178	12	110,492,139	TC	1.6E−07	0.84
16p13.13	CIITA,CLEC16A,SOCS1	rs3862469	16	11101581	TC	1.7E−07	0.82
9q31.1	STX17, NR4A3	rs10124366	9	101,727,524	AG	1.1E−05	0.83

GWAS, genome-wide association studies; OR, odds ratio; SNP, single-nucleotide polymorphism.
Fourteen regions in the genome have demonstrated association with AA by GWAS and replication. For each region, genes of interested are listed, along with the most significant SNP, its location in the genome, *P* value and OR obtained with logistic regression from the combined analysis ($N = 10,796$). *P* values and effect estimates from each analytic stage are in Supplementary Table 3.
P-values in bold exceed the threshold for genome-wide statistical significance.

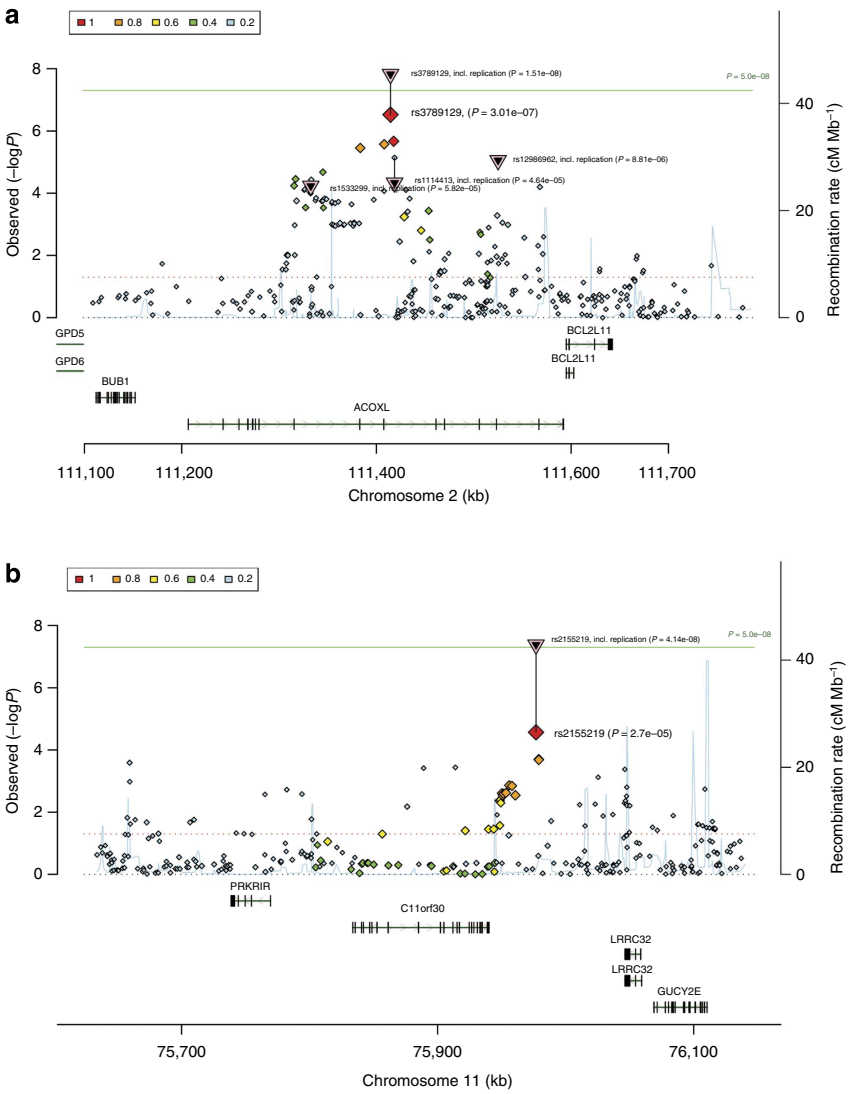


Figure 3 | Detailed map of associated SNPs and gene locations for newly identified loci. Two regions in the genome exceeded statistical significance when the replication data were combined with the meta-analysis results ($N = 10,796$) and analysed with logistic regression. (**a** Chromosome 2q13 includes *ACOXL* and *BCL2L11* (*BIM*). (**b**) Chromosome 12q24.12 included *C11orf30* and *LRCC32* (*GARP*).

their increased suppressive activity, and GARP knockdown reduces their suppressive activity[24,27]. Not surprisingly, we found that GARP is highly expressed in whole PBMCs, consistent with previous reports. Unexpectedly, we also detected a strong signal for GARP in plucked control SHFs, a site not previously described and perhaps atypical for the expression of a Treg protein (Fig. 4). This suggests a potential role for GARP in hair biology and implicates significance of its disruption in AA at the site of pathology for the phenotype.

A third novel region on chromosome 12q24.12 that achieved suggestive evidence for association ($P = 1.3 \times 10^{-7}$) is of interest, because several genes harboured within the region functionally align with other associated genes. This region has demonstrated associations with multiple autoimmune diseases, and contains 10 genes including SH2B adaptor protein 3 (*SH2B3*) and aldehyde dehydrogenase 2 family (*ALDH2*). *SH2B3*, also known as LNK, is a key negative regulator of cytokine signalling via receptor tyrosine kinases and JAK signalling. Two missense variants that confer increased cytokine production and enhanced signalling have been reported in the literature (rs3184504 and rs72650673)[28,29]. We genotyped these two functional polymorphisms in a sample of 96 chromosomes from AA patients carrying the AA-associated risk allele (rs653178*G), and found that 93 chromosomes also carried the rs3184504*T risk allele ($f = 0.97$). This increase in allele frequency over what would be expected among a sample of European Americans ($f = 0.51$ in Exome Variant Server ESP6500) indicates that there is strong linkage disequilibrium (LD) between the AA-associated risk variant and this functional polymorphism, and suggests that R262W could be contributing to AA pathogenesis. The rs72650673*A allele was not found in our sample, consistent with its low allele frequency among European Americans ($f = 0.002$ in Exome Variant Server ESP6500). We found by RT–PCR analysis that LNK is highly expressed in whole PBMCs, as well as T cells, natural killer, B cells and MCs, but did not detect expression in plucked SHFs (Fig. 4). *ALDH2* has been identified as a citrullinated antigen in rheumatoid arthritis patients[30]. This gene has been studied in the skin, with expression localizing to epidermis, sebaceous glands and hair follicles, where it is hypothesized to reduce the accumulation of oxidative stress-induced aldehydes[31].

The strong association with HLA class II in AA points towards the involvement of CD4+ T cells in the pathogenesis.

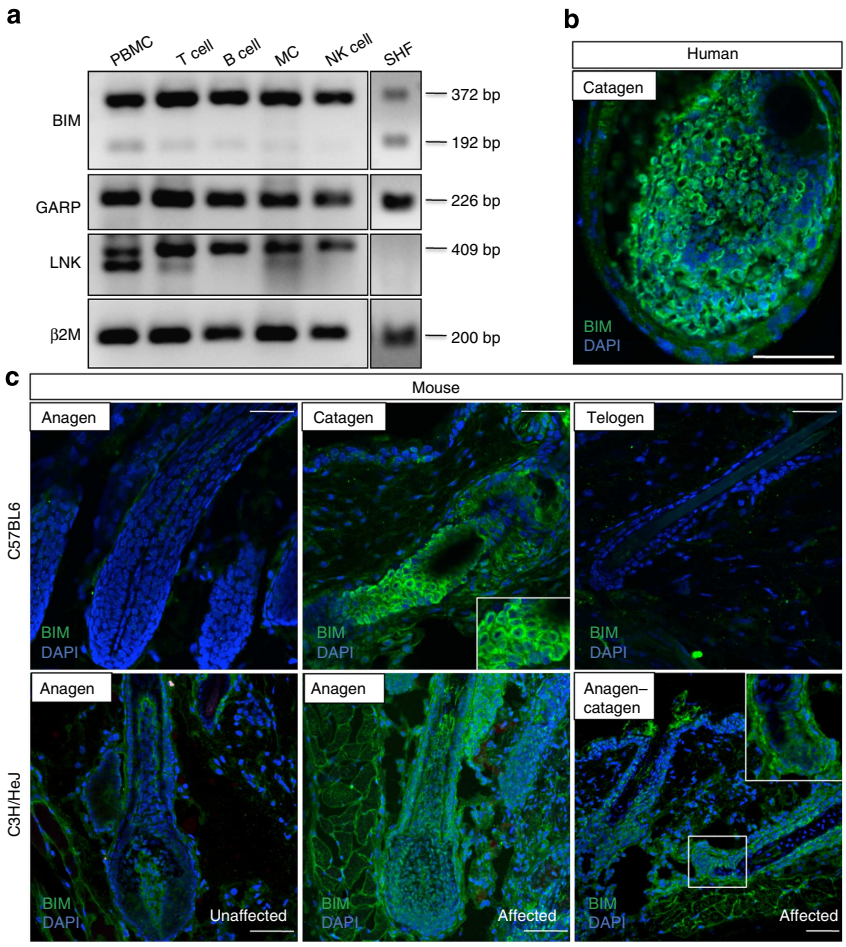


Figure 4 | Characterization BIM, GARP and LNK expression. (**a**) *BIM*, *GARP* and *LNK* expression in human immune cells and hair follicles. RT-PCR was performed on *BIM*, *GARP* and *LNK* to determine gene expression in T cells, natural killer cells (NK), B cells, monocytes (MC), peripheral blood mononuclear cells (PBMCs) and scalp hair follicles (SHF). β2M PCR was used as a loading control for each complementary DNA. Two splice variants are observed for *BIM* expression: BIM-S (192 bp) and BIM-L (372 bp). Expected amplicon sizes: *BIM* (372 bp and 192 bp), *GARP* (226 bp) and *LNK* (409 bp). (**b**) Immunofluorescence staining in human hair follicles reveals that BIM is highly expressed in matrix cells of the catagen hair bulb. (**c**) In healthy hair follicles from the C67BL6 mouse strain, BIM is strongly expressed in the apoptosing strand of mouse catagen hair follicles and is absent from anagen and telogen hair follicles. Immunofluorescence staining on affected and unaffected skin from the AA mouse model strain, C3H/HeJ, reveals that BIM expression levels and the localization pattern are aberrant in affected hair follicles compared with unaffected C3H/HeJ and C57BL6 hair follicles. DAPI, 4′,6-diamidino-2-phenylindole.

The importance of this T-cell subset is supported by the presence of perifollicular CD4+ T cells in addition to intrabulbar CD8+ T cells in human AA. Finally, we analysed these data in the context of other autoimmune diseases by performing CPMA using the 107 SNPs used in the original description of the analysis and data from seven other autoimmune diseases[32]. We found 50 SNPs that are associated at or near significance across two or more diseases, and clustering of these according to their association with disease identified five groups (Fig. 5a,b). Protein–protein interaction graphs of each SNP cluster demonstrate that for several of the groups, the proteins coded from these regions interact either directly or via an intermediary to a significant degree (Fig. 5c). Integration of AA into the CPMA demonstrates mechanistic alignment with coeliac disease, type 1 diabetes, rheumatoid arthritis, multiple sclerosis and Crohn's disease, with most of the overlap coming from the first cluster of genes.

Collectively, this meta-analysis for the first time resolves HLA associations and demonstrates the pivotal aetiological role of HLA-DR. Furthermore, the identification of specific residues in HLA-DRβ1 that are over-represented among AA patients will allow us to better model peptide class II MHC interactions and thus predict autoantigens in AA. Associations outside HLA provide new evidence for the importance of Treg maintenance and immune response pathways in AA. Furthermore, autophagy and apoptosis are emerging as processes of aetiologic importance in AA. These insights will allow us to better understand the molecular taxonomy of autoimmune diseases and the alignment of AA within this class of disorders. Importantly, as GWAS help to resolve disease mechanisms and identify pathogenic pathways perturbed in AA and autoimmunity in general, these approaches advance the field towards precision medicine in autoimmunity.

Methods

Patient population. All participating studies were reviewed and approved by the Institutional Review Boards and ethics committees at Columbia University; MD Anderson Cancer Center; University of Minnesota, Minneapolis; University of Colorado, Denver; University of California, San Francisco; and the Universities of Bonn, Düsseldorf, Münster, Berlin, Hamburg and Munich, Germany; and Antwerp, Belgium, and were conducted in accordance with the Declaration of Helsinki Principles. All study subjects provided written informed consent.

For the previously published GWAS and the ImmunoChip samples, cases were ascertained through the National Alopecia Areata Registry, which recruits patients in

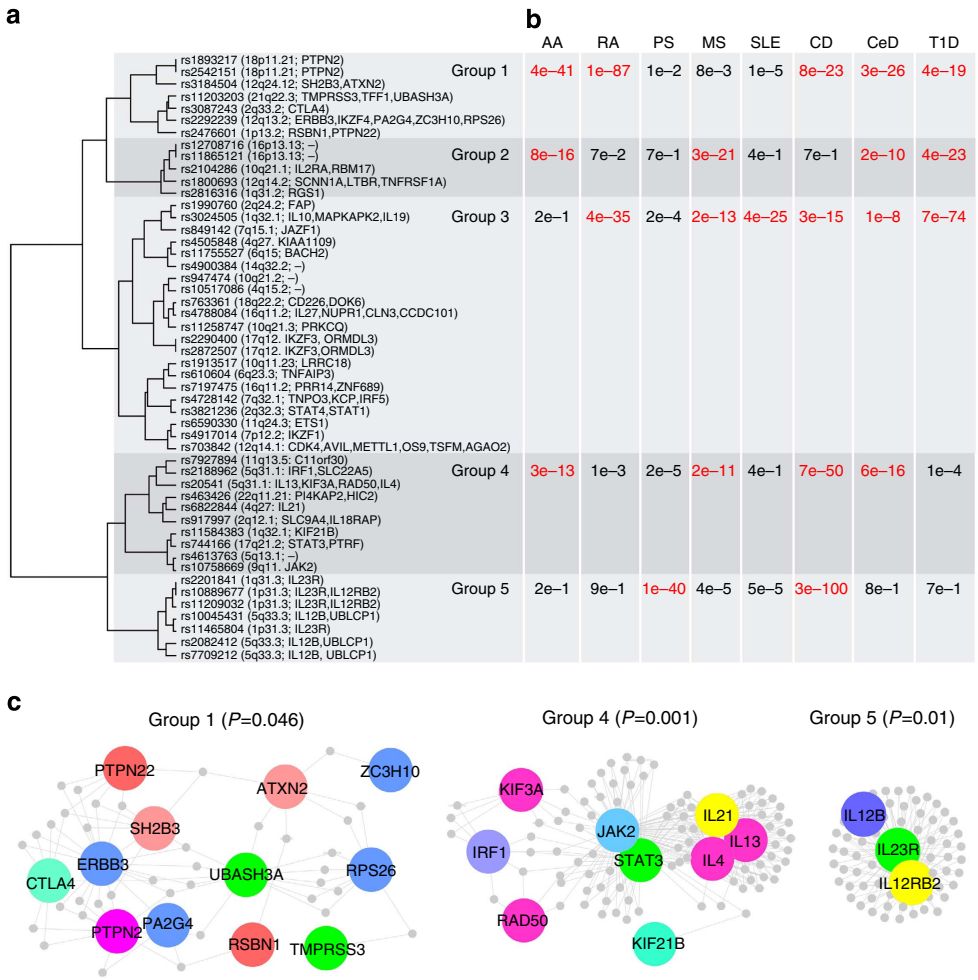


Figure 5 | Cross-phenotype meta-analysis revealing functional clusters of genes and proteins indicating patterns of shared disease mechanisms across autoimmune diseases. (**a**) Fifty SNPs with evidence of association to more than one autoimmune disease ($P_{cpma}<0.01$) are clustered by association with disease. (**b**) For each cluster and disease pair, a cumulative association statistic was calculated using Fisher's omnibus test to combine *P* values. The varying pattern of disease association for each cluster suggests each group represents a distinct co-morbid mechanism. (**c**) Proteins coded within a 100-Mb window centred on each SNP within each cluster are depicted in protein–protein interaction maps. Three of the five clusters have significant protein inter-connectivity (permuted $P<0.05$).

the US through five clinical sites and confirms diagnosis. Control data for these samples were obtained from publically available data. There were two sources of control data for the GWAS[4]. First, a data set was obtained from subjects enrolled in the New York Cancer Project[33] and genotyped as part of previous studies[34]. Second, a data set was obtained from the Cancer Genetic Markers of Susceptibility (CGEMS) breast[35] and prostate[36] cancer studies (http://cgems.cancer.gov/data/). The ImmunoChip control data were obtained from the NIDDK inflammatory bowel disease genetics consortium (http://medicine.yale.edu/intmed/ibdgc/index.aspx). The cases and controls for the previously unreported GWAS were ascertained from outpatient clinics, private dermatology practices and via AA self-support groups in Belgium and Germany. Inclusion criteria followed published guidelines[37], and additionally included diagnosis by a trained and experienced clinician. European GWAS controls included population-based controls established within the National Genome Research Network for use as universal controls (PopGen[38], KORA[39]), and the Heinz Nixdorf Recall Study[40] and an additional population-based cohort from the Heinz Nixdorf Recall[40] study. Sample counts by ascertainment location and genotyping platform are summarized in Supplementary Table 1.

Genotyping. DNA was extracted from peripheral blood leukocytes by salting out with saturated NaCl solution according to standard methods, or by using a Chemagic Magnetic Separation Module I (Chemagen, Baesweiler, Germany), in accordance with the manufacturer's instructions. Whole-genome genotyping for the meta-analayis was performed on either Illumina HumanHap550 BeadChip or the Illumina Omni express, as detailed in Supplementary Table 1. The replication cohort was genotyped with the Illumina Immunochip. In addition, SNPs selected for replication, which were not present on the Immunochip, were genotyped on the MassArray system using a Sequenom Compact MALDI-TOF device and iPLEX Gold reagents (Sequenom, San Diego, CA) in multiplex reactions. Primer sequences and standard assay conditions are available upon request. Standard quality-control (QC) metrics were employed, removing data for SNPs and samples with call rates less than 90. Additional genotyping was performed by standard PCR-based techniques.

Discovery QC and imputation. Technical QC was performed with QC conducted on each data set separately using a common approach. The following QC parameters were applied: (i) missing rate per SNP <0.05 (before sample removal below), (ii) missing rate per individual <0.02, (iii) heterozygosity per individual ± 0.2 (Fhet, statistic, standardized values in plink), (iv) missing rate per SNP <0.02 (after sample removal above), (iv) missing rate per SNP difference in cases and controls <0.02, (vi) Hardy–Weinberg equilibrium (in controls) $P<10^{-6}$, (vii) Hardy–Weinberg equilibrium (in cases) $P<10^{-10}$. Study sample sizes varied between 1,200 and 3,000 individuals. The number of SNPs per study after QC varied between 420,000 and 640,000. On average, the QC processes excluded 44 individuals per study (with a range of 13–107 individuals) and 13,000 SNPs per study (with a range of 3,000–20,000 SNPs). After QC, the GWAS data sets together comprised 2,489 cases and 5,287 controls and, for the next steps of the 'genetic QC' analysis, a set of 221,784 SNPs common to all platforms and successfully genotyped in each GWAS sample was extracted. These SNPs were then further pruned to remove LD (leaving no pairs with $r^2>0.05$) and lower-frequency SNPs (minor allele frequency <0.05), leaving 64,099 SNPs suitable for robust relatedness testing and population structure analysis (see below). Imputation of untyped SNPs was performed within each study in batches of 300 individuals. These batches were randomly drawn to keep the same case-control ratio as in the total sample from that study. We used Beagle 3.13. Imputation was performed with CEU + TSI HapMap phase 3 data (UCSC hg18/

NCBI 36) using a chunk size of 10 Mb with 410 phased haplotypes comprising 1,233,578 SNPs, using default parameters. Lambda-GC was carefully monitored before and after imputation.

Genetic QC included relatedness testing and principal component analyses based on 64,099 LD-independent SNPs, present on all platforms in this study. Relatedness testing was done with PLINK[41], reporting pairs with genome identity (pi-hat) >0.9 as 'identical samples' and with pi-hat >0.2 as being closely related. After random shuffling, one individual from each related pair was excluded from downstream analysis. From groups with multiple related pairs (for example, a family), only one individual was kept.

Discovery principal component analysis. Principal component estimation was done with the same collection of SNPs on the non-related subset of individuals. We estimated the first 20 principal components and tested each of them for phenotype association (using logistic regression with study indicator variables included as covariates) and evaluated their impact on the genome-wide test statistics using Lamda-GC (the genomic control inflation factor based on the median χ^2) after genome-wide association of the specified principal component. On the basis of this, we decided to include principal components 1, 2, 3, 4, 5 and 6 for downstream analysis as associated covariates.

Association analysis. The primary scan had a total of 2,332 QC+ cases and 5,233 QC+ controls. Association testing was carried out in PLINK[41], using the dosage data from the imputation and using the first six principal components as covariates, chosen as described above from the first 20 principal components. The four data sets had Lamda-GC values of 1.025, 1.041, 1.040 and 1.042.

A fixed-effect meta-analysis was performed using ORs and s.e. The final meta-analysis had a Lambda-GC of 1.076.

MHC imputation and association analysis. Starting from the genotyped SNPs, we imputed classical HLA alleles and corresponding polymorphic amino acids within classical HLA proteins, following a recently described protocol[8]. In total, 8,271 variants including SNPs, amino acids and classical HLA alleles at 2- and 4-digit resolution were tested for association in the MHC region. Logistic regression was applied for testing for association, including the six first principal components and a dummy variable for the cohort identification of each sample. Single-test association was performed for all types of variants, as well as omnibus tests for the amino acids. The conditional analyses were performed using the initial principal component analysis covariates, the dummy variable for the cohort, and adjusting for the top and conditional hits.

Analysis of immunochip data. To identify the sample ethnicities in the case cohort, we calculated the principal components (PC) 1 and 2 using Hapmap 3, and mapped the case samples to the principal components. Samples that do not cluster with the Hapmap3 European samples were discarded. The NIDDK control cohort has been cleaned for sample ethnicity and is thus not necessary to undergo such procedure. For both case and control cohort, we then removed samples that have missing genotype rate $>2\%$ and variants that have missing genotype rate $>2\%$ for each cohort. For the control cohort, we additionally removed variants that fail the Hardy–Weinberg equilibrium test (P value $<1\times10^{-5}$). The case and the control cohorts were then merged, and we removed variants that have different missingness in cases and in controls (P value $<1\times10^{-5}$) and samples that have high heterozygosity (>3 s.d.). We used PLINK to infer the genetic relationships between the remaining samples and we randomly removed one of the samples for each related sample pair (PI_HAT >0.5)[41]. The final cleaned data set has 318 cases and 1,688 controls. We tested the association between the genetic variants and AA using PLINK with the principal components 1–5 as covariates.

RT-PCR and immunofluorescence staining. RT–PCR was performed on total RNA extracted from plucked human SHFs and T cells, natural killer cells, B cells, MCs, PBMCs that were isolated by FACS from whole blood. The primer sequences of the *BIM*, *GARP*, *LNK* and *β2M* genes are as follows: BIM F: 5′-TAAGTTC TGAGTGTGACCGAGA-3′, R: 3′-CCATTGCACTGAGATAGTGGTTG-5′; GARP F: 5′-CGCTCCCGAGACTCATCTAC-3′, R: 5′-AGGTGCTCAAGAAAG CTGTCG-3′; LNK F: 5′-GTGGGGAATACGTGCTCACTT-3′, R: 5′-TGTCCA CGACCGAGGGAAA-3′; β2M F: 5′-GAGGCTATCCAGCGTACTCCA-3′, R: 5′-CGGCAGGCATACTCATCTTTT-3′. Full blots are provided as a Supplementary Figure 1.

Immunofluorescence staining was performed on human and mouse samples. Human skin and hair follicle sections were obtained from control occipital scalp biopsies discarded during surgery and considered to be non-human subject research under 45 CFR Part 46 (Institutional Review Board exempted). Mouse anagen skin (day 30), catagen skin (day 42) and telogen skin (day 50) was collected from the dorsal region and embedded in optimal cutting temperature for sectioning (10 μm). Slides were fixed with 50% MeOH/50% acetone for 10 min at 220 μC, washed with 1 × phosphate-buffered saline and then blocked with 2% fish skin gelatin (Sigma Aldrich, MO, USA). The α-BIM antibody (Rabbit) (Biorbyt, orb10190) was used at a concentration of 1:200 for both mouse and human studies. Slides were washed, incubated with the Alexa Fluor 488 donkey anti-rabbit IgG (Molecular Probes, Invitrogen) secondary antibody (1:800 in 1 × phosphate-buffered saline), mounted with VECTASHIELD mounting medium with DAPI (Vector Laboratories, Burlingame, CA, USA) and imaged using a LSM 5 laser-scanning Axio Observer Z1 confocal microscope (Carl Zeiss).

Cross-phenotype meta-analysis. The protocol for this analysis was adopted from previously described work on cross-phenotype meta-analysis[32]. Using Dataset S1 from the Supplementary Data of the original analysis, we added in AA association P values for the 107 SNPs originally investigated. For SNPs with no P value, we used Broad Institutes SNAP to find a proxy SNP with $r^2>0.9$ (http://www.broadinstitute.org/mpg/snap/). The CPMA statistic was then recalculated for all 107 SNPs using a code (found here: http://coruscant.itmat.upenn.edu/software.html) implementing the original description. All SNPs meeting a significance threshold of <0.01 were included in clustering analysis. Clustering was done by first binning P values into four groups based on magnitude and clustered using the Cluster package. Cumulative association statistics were calculated in R using Fisher's omnibus test. Protein–protein interaction analysis was performed using DAPPLE, a publicly available web application (http://www.broadinstitute.org/mpg/dapple). Settings for DAPPLE included the 1000 genome assembly, 5,000 permutations and 50-kb regulatory region upstream and downstream. All statistical analyses were performed in R version 3.0.1.

References

1. Safavi, K. H., Muller, S. A., Suman, V. J., Moshell, A. N. & Melton, 3rd L. J. Incidence of alopecia areata in Olmsted County, Minnesota, 1975 through 1989. *Mayo Clin. Proc.* **70,** 628–633 (1995).
2. Delamere, F. M., Sladden, M. M., Dobbins, H. M. & Leonardi-Bee, J. Interventions for alopecia areata. *Cochrane Database Syst. Rev.* CD004413 (2008).
3. Lebwohl, M. *Treatment of Skin Disease: Comprehensive Therapeutic Strategies* 2nd edn (Mosby/Elsevier, 2006).
4. Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466,** 113–117 (2010).
5. Jagielska, D. *et al.* Follow-up study of the first genome-wide association scan in alopecia areata: IL13 and KIAA0350 as susceptibility loci supported with genome-wide significance. *J. Invest. Dermatol.* **132,** 2192–2197 (2012).
6. John, K. K. *et al.* Genetic variants in CTLA4 are strongly associated with alopecia areata. *J. Invest. Dermatol.* **131,** 1169–1172 (2011).
7. Redler, S. *et al.* Investigation of selected cytokine genes suggests that IL2RA and the TNF/LTA locus are risk factors for severe alopecia areata. *Br. J. Dermatol.* **167,** 1360–1365 (2012).
8. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8,** e64683 (2013).
9. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501,** 506–511 (2013).
10. Entz, P. *et al.* Investigation of the HLA-DRB1 locus in alopecia areata. *Eur. J. Dermatol.* **16,** 363–367 (2006).
11. Colombe, B. W., Lou, C. D. & Price, V. H. The genetic basis of alopecia areata: HLA associations with patchy alopecia areata versus alopecia totalis and alopecia universalis. *J. Investig. Dermatol. Symp. Proc.* **4,** 216–219 (1999).
12. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44,** 291–296 (2012).
13. Melum, E. *et al.* Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat. Genet.* **43,** 17–19 (2010).
14. Bouillet, P., Cory, S., Zhang, L. C., Strasser, A. & Adams, J. M. Degenerative disorders caused by Bcl-2 deficiency prevented by loss of its BH3-only antagonist Bim. *Dev. Cell* **1,** 645–653 (2001).
15. Santin, I. *et al.* PTPN2, a candidate gene for type 1 diabetes, modulates pancreatic beta-cell apoptosis via regulation of the BH3-only protein Bim. *Diabetes* **60,** 3279–3288 (2011).
16. Luo, S. Q. & Rubinsztein, D. C. BCL2L11/BIM: a novel molecular link between autophagy and apoptosis. *Autophagy* **9,** 104–105 (2013).
17. Hamasaki, M. *et al.* Autophagosomes form at ER-mitochondria contact sites. *Nature* **495,** 389–393 (2013).
18. Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151,** 1256–1269 (2012).
19. O'Reilly, L. A. *et al.* The proapoptotic BH3-only protein bim is expressed in hematopoietic, epithelial, neuronal, and germ cells. *Am. J. Pathol.* **157,** 449–461 (2000).
20. Nogueira, T. C. *et al.* GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. *PLoS Genet.* **9,** e1003532 (2013).
21. Jiang, C. C. *et al.* Apoptosis of human melanoma cells induced by inhibition of B-RAFV600E involves preferential splicing of bimS. *Cell Death Dis.* **1,** e69 (2010).

22. Tobin, D. J. Aging of the hair follicle pigmentation system. *Int. J. Trichol.* **1,** 83–93 (2009).
23. Battaglia, M. & Roncarolo, M. G. The Tregs' world according to GARP. *Eur. J. Immunol.* **39,** 3296–3300 (2009).
24. Tran, D. Q. *et al.* GARP (LRRC32) is essential for the surface expression of latent TGF-beta on platelets and activated FOXP3+ regulatory T cells. *Proc. Natl Acad. Sci. USA* **106,** 13445–13450 (2009).
25. Stockis, J., Colau, D., Coulie, P. G. & Lucas, S. Membrane protein GARP is a receptor for latent TGF-beta on the surface of activated human Treg. *Eur. J. Immunol.* **39,** 3315–3322 (2009).
26. Wang, R. *et al.* GARP regulates the bioavailability and activation of TGFbeta. *Mol. Biol. Cell* **23,** 1129–1139 (2012).
27. Wang, R. *et al.* Expression of GARP selectively identifies activated human FOXP3+ regulatory T cells. *Proc. Natl Acad. Sci. USA* **106,** 13439–13444 (2009).
28. Zhernakova, A. *et al.* Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am. J. Hum. Genet.* **86,** 970–977 (2010).
29. McMullin, M. F., Wu, C., Percy, M. J. & Tong, W. A nonsynonymous LNK polymorphism associated with idiopathic erythrocytosis. *Am. J. Hematol.* **86,** 962–964 (2011).
30. Wegner, N. *et al.* Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. *Immunol. Rev.* **233,** 34–54 (2010).
31. Cheung, C., Davies, N. G., Hoog, J. O., Hotchkiss, S. A. M. & Pease, C. K. S. Species variations in cutaneous alcohol dehydrogenases and aldehyde dehydrogenases may impact on toxicological assessments of alcohols and aldehydes. *Toxicology* **184,** 97–112 (2003).
32. Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7,** e1002254 (2011).
33. Mitchell, M. K., Gregersen, P. K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health* **81,** 301–310 (2004).
34. Plenge, R. M. *et al.* TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *New Engl. J. Med.* **357,** 1199–1209 (2007).
35. Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39,** 870–874 (2007).
36. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39,** 645–649 (2007).
37. Olsen, E. A. *et al.* Alopecia areata investigational assessment guidelines--Part II. National Alopecia Areata Foundation. *J. Am. Acad. Dermatol.* **51,** 440–447 (2004).
38. Krawczak, M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* **9,** 55–61 (2006).
39. Wichmann, H. E., Gieger, C., Illig, T. & Group, M. K. S. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67**(Suppl 1): S26–S30 (2005).
40. Schmermund, A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am. Heart J.* **144,** 212–218 (2002).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).

Acknowledgements

We are deeply indebted to the many patients and their family members who participated in this study. We are thankful to Drs Lawrence Shapiro and Tatyana Gindin for guidance on the structural effects of HLA associations. We thank Spandon V. Shah, Holly Jiang, Matthew Ding and Steven Chiu for their assistance and contributions in isolating the immune cells by FACS and Jane E. Cerise for her bioinformatic expertise. We are most grateful for the support of the National Alopecia Areata Foundation for funding the initial studies, and Ms Vicki Kalabokes and her staff at NAAF for their efforts on our behalf. The US patient cohort was collected and maintained by the National Alopecia Areata Registry (N01-AR62279) (to M.D.). Some controls were drawn from the Heinz Nixdorf Recall Study cohort, which was established with the support of the Heinz Nixdorf Foundation. R.C.B. and M.M.N. are members, T.B. is an associate member, of the DFG-funded Excellence Cluster ImmunoSensation. R.C.B. is a recipient of a Heisenberg Professorship of the German Research Foundation (DFG). This work was supported in part by the DFG grant BE 2346/5-1, as well as by local funding (BONFOR) to R.C.B., Vernieuwingsimpuls VIDI Award from the Netherlands Organization for Scientific Research for project 016.126.354 to P.I.W.d.B., and USPHS NIH/NIAMS grants R01AR52579 and R01AR56016 (to A.M.C.).

Author contributions

R.C.B., M.M.N., S. Redler, S.M., M.D., V.H.P., M.H., D.N. and J.M.-W. participated in phenotyping, diagnosis and access to patient samples. L.P. performed technical aspects in preparation of samples for genotyping and data transfer. L.B. and S.H. performed genotyping. H.H., L.P., S. Ripke, T.Y., T.B., A.M. and P.I.W.d.B. performed statistical analyses. G.M.D. performed immunofluorescence staining and RT–PCR experiments. A.L. and P.K.G. provided control samples and performed genotyping, as well as insight into autoimmune diseases. C.I.A. provided additional statistical advice and control samples. L.P. contributed to study design, interpretation of results and drafting of the manuscript. R.C.B., M.M.N., M.D., D.A. and A.M.C. provided conceptual guidance to the project. A.d.J., R.C., A.M. and P.I.W.d.B. provided expertise on interpreting HLA results. A.M.C. additionally supplied input into the functional significance of candidate genes, supervision of lab personnel, management of collaborations, preparation of the manuscript and all reporting requirements for granting agencies.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Betz, R. C. *et al.* Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat. Commun.* 6:5966 doi: 10.1038/ncomms6966 (2015).

Chapter IV.3

Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition

(Manuscript #4)

nature medicine

Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition

Luzhou Xing[1,7], Zhenpeng Dai[2,7], Ali Jabbari[2,7], Jane E Cerise[2,3], Claire A Higgins[2], Weijuan Gong[2], Annemieke de Jong[2], Sivan Harel[2], Gina M DeStefano[2,4], Lisa Rothman[2], Pallavi Singh[2], Lynn Petukhova[2], Julian Mackay-Wiggan[2], Angela M Christiano[2,5,8] & Raphael Clynes[1,2,6,8]

Alopecia areata (AA) is a common autoimmune disease resulting from damage of the hair follicle by T cells. The immune pathways required for autoreactive T cell activation in AA are not defined limiting clinical development of rational targeted therapies[1]. Genome-wide association studies (GWAS)[2] implicated ligands for the NKG2D receptor (product of the *KLRK1* gene) in disease pathogenesis. Here, we show that cytotoxic CD8$^+$NKG2D$^+$ T cells are both necessary and sufficient for the induction of AA in mouse models of disease. Global transcriptional profiling of mouse and human AA skin revealed gene expression signatures indicative of cytotoxic T cell infiltration, an interferon-γ (IFN-γ) response and upregulation of several γ-chain (γ_c) cytokines known to promote the activation and survival of IFN-γ–producing CD8$^+$NKG2D$^+$ effector T cells. Therapeutically, antibody-mediated blockade of IFN-γ, interleukin-2 (IL-2) or interleukin-15 receptor β (IL-15Rβ) prevented disease development, reducing the accumulation of CD8$^+$NKG2D$^+$ T cells in the skin and the dermal IFN response in a mouse model of AA. Systemically administered pharmacological inhibitors of Janus kinase (JAK) family protein tyrosine kinases, downstream effectors of the IFN-γ and γ_c cytokine receptors, eliminated the IFN signature and prevented the development of AA, while topical administration promoted hair regrowth and reversed established disease. Notably, three patients treated with oral ruxolitinib, an inhibitor of JAK1 and JAK2, achieved near-complete hair regrowth within 5 months of treatment, suggesting the potential clinical utility of JAK inhibition in human AA.

Alopecia areata is a T cell–mediated autoimmune disease characterized phenotypically by hair loss and, histologically, by infiltrating T cells surrounding the hair follicle bulb (reviewed in ref. 1). Previous studies have shown that transfer of total T cells (but not B cells or sera) can cause the disease in human xenograft models[3], as well as in C3H/HeJ mice[4], a mouse strain that develops spontaneous AA with considerable similarity to human AA. Broad-acting intralesional steroids are the most commonly used therapy for AA, with varying success. Progress in developing effective, rationally targeted therapies has been limited by our lack of mechanistic understanding of the underlying key T cell inflammatory pathways in AA.

We[2] and others[5] have previously identified a cytotoxic subset of CD8$^+$NKG2D$^+$ T cells within the infiltrate surrounding human AA hair follicles, as well as concomitant upregulation in the follicle itself of the 'danger signals' ULBP3 (ref. 2) and MICA[5], two NKG2D ligands (NKG2DLs) whose importance in disease pathogenesis has also been suggested by genome-wide association studies[2].

To determine the contribution of CD8$^+$NKG2D$^+$ T cells to AA pathogenesis, we used the C3H/HeJ mouse model[6], which spontaneously develops alopecia and recapitulates many pathologic features of human AA[7]. In lesional skin biopsies from alopecic mice, CD8$^+$NKG2D$^+$ T cells infiltrate the epithelial layers of the hair follicle, which overexpress the NKG2DLs, H60 and Rae-1, analogous to what has been observed in skin biopsies of human AA[2] (**Fig. 1a**,**b** and **Supplementary Fig. 1a**,**b**). Flow cytometric analysis of the CD45$^+$ leukocyte population in the skin revealed a marked increased number of CD8$^+$NKG2D$^+$ T cells in the skin of diseased C3H/HeJ mice, in conjunction with cutaneous lymphadenopathy and increased total cellularity, as compared with disease-free C3H/HeJ mice (**Fig. 1c**,**d**). Other cell types, including CD4$^+$ T cells[4] and mast cells[8], were present in much smaller numbers (**Supplementary Fig. 1c** and data not shown).

The immunophenotype of the skin-infiltrating CD8$^+$ T cells in mice with AA was similar to that of the CD8$^+$NKG2D$^+$ population found in the cutaneous lymph nodes: CD8αβ$^+$ effector memory T cells (T_{EM}, CD8^{hi}CD44^{hi}CD62L^{low}CD103$^+$) bearing several natural killer (NK) immunoreceptors, including CD49b and NKG2A, NKG2C and NKG2E (**Fig. 1e** and **Supplementary Fig. 1d**). These CD8$^+$ T_{EM} cells expressed high levels of IFN-γ and exhibited NKG2D-dependent cytotoxicity against *ex vivo*–expanded syngeneic dermal sheath target cells (**Fig. 1f**). Gene expression analysis of the CD8$^+$NKG2D$^+$ T cells isolated from alopecic C3H/HeJ lymph node cells using

[1]Department of Pathology, Columbia University, New York, New York, USA. [2]Department of Dermatology, Columbia University, New York, New York, USA. [3]Department of Psychiatry, Columbia University, New York, New York, USA. [4]Department of Epidemiology, Columbia University, New York, New York, USA. [5]Department of Genetics and Development, Columbia University, New York, New York, USA. [6]Department of Medicine, Columbia University, New York, New York, USA. [7]These authors contributed equally to this work. [8]These authors jointly directed this work. Correspondence should be addressed to A.M.C. (amc65@columbia.edu) or R.C. (rc645@columbia.edu).

Received 22 December 2013; accepted 30 June 2014; published online 17 August 2014; doi:10.1038/nm.3645

RNA-seq demonstrated a transcriptional profile characteristic of effector cytotoxic T lymphocytes (CTLs)[9,10] and identified several additional NK-specific transcripts (**Supplementary Table 1** and **Supplementary Fig. 2**).

We next evaluated the requirement of these $CD8^+$ T_{EM} cells in disease pathogenesis. Transfer of cytotoxic $CD8^+NKG2D^+$ cells or total lymph node cells from diseased mice induced AA in all five healthy C3H/HeJ recipients by 14 weeks after transfer, whereas lymph node cell populations depleted of $NKG2D^+$ cells were unable to transfer disease (**Fig. 1g**). Thus, $CD8^+NKG2D^+$ T cells are the dominant cell type in the dermal infiltrate and are necessary and sufficient for T cell–mediated transfer of AA.

To characterize the transcriptional profile of AA lesional skin from C3H/HeJ mice as well as human AA, we performed Affymetrix microarray analyses to identify differentially expressed genes in skin between individuals with AA and skin from control individuals without disease (**Supplementary Fig. 3** and **Supplementary Tables 2** and **3**) Three gene expression signatures were identified in lesional skin: IFN response genes, such as those encoding the IFN-inducible chemokines CXCL-9, CXCL-10 and CXCL-11 (**Supplementary Figs. 3** and **4**), several key CTL-specific transcripts, such as those encoding CD8A and granzymes A and B, and γ_c cytokines and their receptors, such as the transcripts for interleukin-2 (IL-2) and IL-15, in both human and mouse AA skin. As IL-2Rα was previously shown

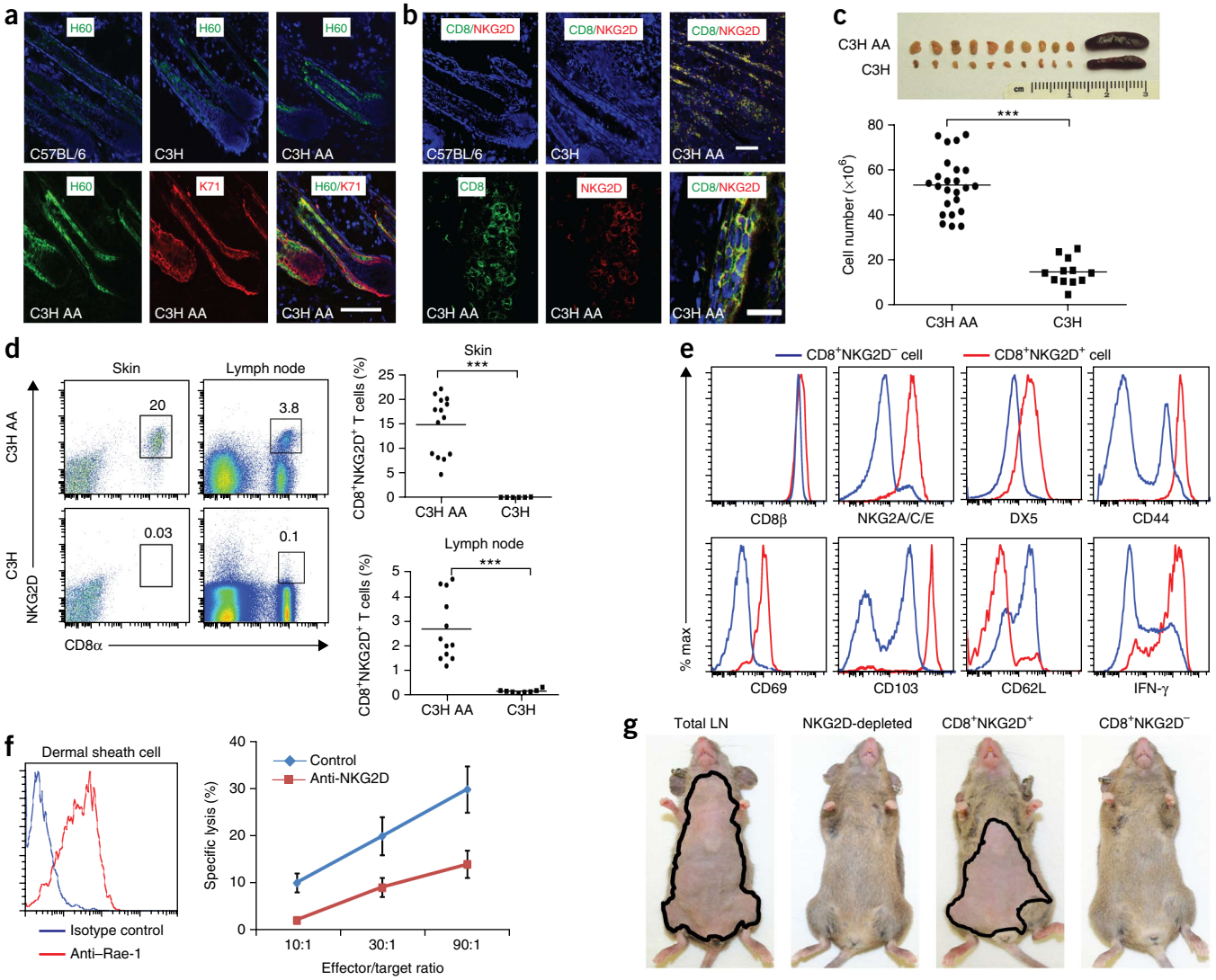


Figure 1 $CD8^+NKG2D^+$ cytotoxic T lymphocytes accumulate in the skin and are necessary and sufficient to induce disease in AA mice. (**a**) Immunofluorescence staining of NKG2D ligand (H60) in the hair follicle inner root sheath (marked by K71). Scale bar, 100 μm. (**b**) $CD8^+NKG2D^+$ cells in hair follicles of C57BL/6, healthy C3H/HeJ and C3H/HeJ AA mice. Top scale bar, 100 μm; bottom scale bar, 50 μm. (**c**) Cutaneous lymphadenopathy and hypercellularity in C3H/HeJ AA mice. (**d**) Frequency (number shown above boxed area) of $CD8^+NKG2D^+$ T cells in the skin and skin-draining lymph nodes in alopecic mice versus ungrafted mice. (**e**) Immunophenotype of $CD8^+NKG2D^+$ T cells in cutaneous lymph nodes of C3H/HeJ alopecic mice. (**f**) Left, Rae-1t–expressing dermal sheath cells grown from C3H/HeJ hair follicles. Right, dose-dependent specific cell lysis induced by $CD8^+NKG2D^+$ T cells isolated from AA mice cutaneous lymph nodes in the presence of blocking anti-NKG2D antibody or isotype control. Effector to target ratio given as indicated. Data are expressed as means ± s.d. (**g**) Hair loss in C3H/HeJ mice injected subcutaneously with total lymph node (LN) cells, $CD8^+NKG2D^+$ T cells alone, $CD8^+NKG2D^-$ T cells or lymph node cells depleted of $NKG2D^+$ (5 mice per group). Mice are representative of two experiments. $^{***}P < 0.001$ (Fisher's exact test). For **c**,**d**,**f**, *n* and number of repeats are detailed in the **Supplementary Methods**.

to be expressed on infiltrating lymphocytes surrounding human AA hair follicles[11], we performed immunofluorescence analysis for both IL-15 and its chaperone receptor IL-15Rα to identify the source of IL-15 in the skin. We detected a marked upregulation of both components in AA hair follicles in both human and mouse AA and found IL-15Rβ expressed on infiltrating CD8$^+$ T cells in humans (**Supplementary Figs. 5** and **6**).

IL-2 and IL-15 are well-known drivers of cytotoxic activity by IFN-γ–producing CD8$^+$ effector T cells and NK cells[12,13] and have been implicated in the induction and/or maintenance of autoreactive

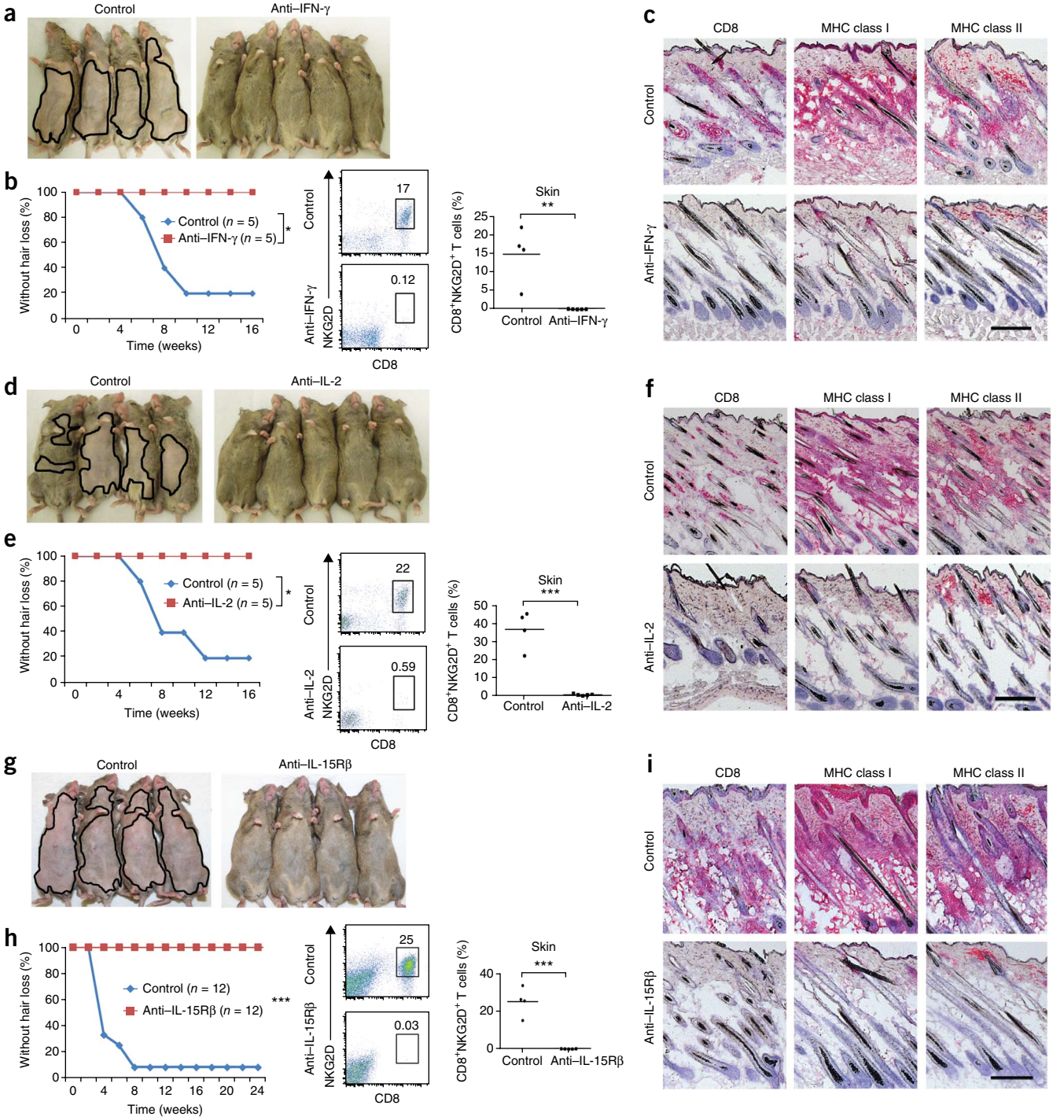


Figure 2 Prevention of AA by blocking antibodies to IFN-γ, IL-2 or IL-15Rβ. C3H/HeJ grafted mice were treated systemically from the time of grafting. (**a**–**h**) AA development in C3H/HeJ grafted mice treated systemically from the time of grafting with antibodies to IFN-γ (**a**,**b**), IL-2 (**d**,**e**) and IL-15Rβ (**g**,**h**). Frequency (number shown above boxed area) of CD8$^+$NKG2D$^+$ T cells in the skin of mice treated with antibodies to IFN-γ (**b**), IL-2 (**e**) and IL-15Rβ (**h**) compared to PBS-treated mice. (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, statistical methods described in the **Supplementary Methods**. Immunohistochemical staining of skin biopsies showing CD8 and MHC class I and II expression in skin of mice treated with isotype control antibody or with antibodies to IFN-γ (**c**), IL-2 (**f**) or IL-15Rβ (**i**). Scale bars, 100 μm. For each experiment, *n* and number of repeats are detailed in the **Supplementary Methods**.

CD8$^+$ T cells[14–16]. To test the efficacy of IFN-γ– and γ_c-targeted therapies *in vivo,* we used the well-established graft model of AA, in which skin grafts from mice with spontaneous AA are transferred onto the backs of unaffected 10-week-old recipient C3H/HeJ mice. In this model, AA develops reliably in 95-100% of grafted recipients within 6–10 weeks[7], allowing us to test interventions aimed at either preventing or reversing disease.

The role of IFN-γ in AA was previously investigated using both knockout studies and administration of IFN-γ, where IFN-γ–deficient mice were resistant and exogenous IFN-γ precipitated disease[17,18]. Administration of neutralizing antibodies to IFN-γ at the time of grafting prevented AA development in grafted recipients and abrogated major histocompatibility complex (MHC) upregulation and CD8$^+$NKG2D$^+$ infiltration in the skin (**Fig. 2a–c**). Likewise, a role for IL-2 in AA pathogenesis was previously established using genetic experiments in which IL-2 haploinsufficiency on the C3H/HeJ background conferred resistance to disease by about 50% using the graft model[19], and this role is supported by our genome-wide association studies in humans[2]. Systemically administered blocking antibodies to either IL-2 (**Fig. 2d–f**) or IL-15Rβ (**Fig. 2g–i**) prevented AA in grafted mice, blocked the accumulation of CD8$^+$NKG2D$^+$ T cells in the skin and abrogated MHC upregulation. However, IL-21 blockade failed to prevent the development of AA in grafted C3H/HeJ mice (**Supplementary Fig. 7**). Notably, none of these blocking antibodies given alone was able to reverse established AA (data not shown).

We next asked whether we could recapitulate the effects of type I cytokine blockade by intervening downstream using small-molecule

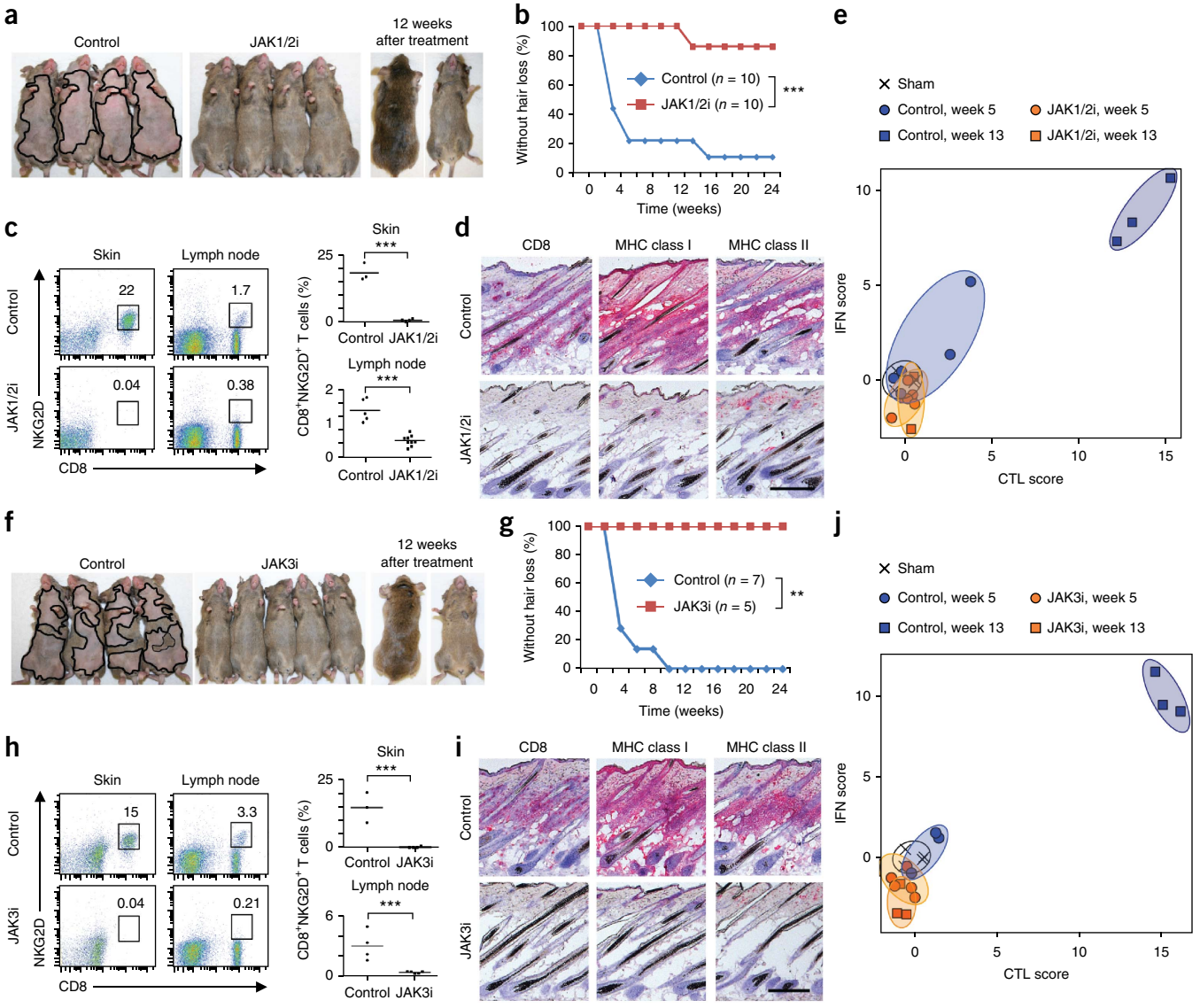


Figure 3 Systemic JAK1/2 or JAK3 inhibition prevents the onset of AA in grafted C3H/HeJ mice. (**a**–**j**) AA development in C3H/HeJ grafted mice treated systemically from the time of grafting with ruxolitinib (JAK1/2i) (**a**,**b**) or tofacitinib (JAK3i) (**f**,**g**) (**$P < 0.01$). Frequency (number shown above boxed area) of CD8$^+$NKG2D$^+$ T cells in skin and cutaneous lymph nodes of mice treated with PBS or with JAK1/2i (**c**) or JAK3i (**h**) (***$P < 0.001$, statistical methods described in **Supplementary Methods**). Immunohistochemical staining of skin biopsies showing CD8 and MHC class I and II expression in skin of mice treated with PBS or with JAK1/2i (**d**) or JAK3i (**i**). ALADIN score of transcriptional analysis from mice treated with PBS or with JAK1/2i (**e**) or JAK3i (**j**), given as $\log_2$ mean expression *Z*-scores as indicated in the **Supplementary Methods**. Hair regrowth after an additional 12 weeks after treatment withdrawal is also shown. (**a**,**f**). Scale bars, 100 μm. For each experiment, *n* and number of repeats are detailed in the **Supplementary Methods**.

inhibitors of JAK kinases, which signal downstream of a wide range of cell surface receptors. In particular, IFN-γ receptors and γ_c family receptors signal through JAK1/2 and JAK1/3, respectively[20]. JAK activation was shown by the presence of phosphorylated signal transducer and activator of transcription (STAT) proteins (pSTAT1, pSTAT3 and to a lesser extent pSTAT5) in human and mouse alopecic hair follicles, but not in normal hair follicles (**Supplementary Fig. 8**). In *in vitro*–cultured dermal sheath cells from C3H/HeJ mice, exogenous IFN-γ increased STAT1 activation, whereas IFN-γ plus TNF-α increased surface IL-15 expression (**Supplementary Fig. 9a**,**b**). Ruxolitinib, a US Food and Drug Administration (FDA)–approved small-molecule inhibitor of the JAK1/2 kinases (JAK selectivity is JAK1 = JAK2>Tyk2>>>JAK3)[21] critical for IFN-γR signaling inhibited these responses (**Supplementary Fig. 9a**,**b**). In cultured CTL effectors from C3H/HeJ mice, the FDA-approved small-molecule JAK3 inhibitor tofacitinib (JAK3>JAK1>>JAK2 selectivity)[22] blocked IL-15–triggered pSTAT5 activation (**Supplementary Fig. 9c**). Tofacitinib also blocked killing of dermal sheath cells (**Supplementary Fig. 9d**) and IL-15-induced upregulation of granzyme B and IFN-γ expression (**Supplementary Fig. 9e**).

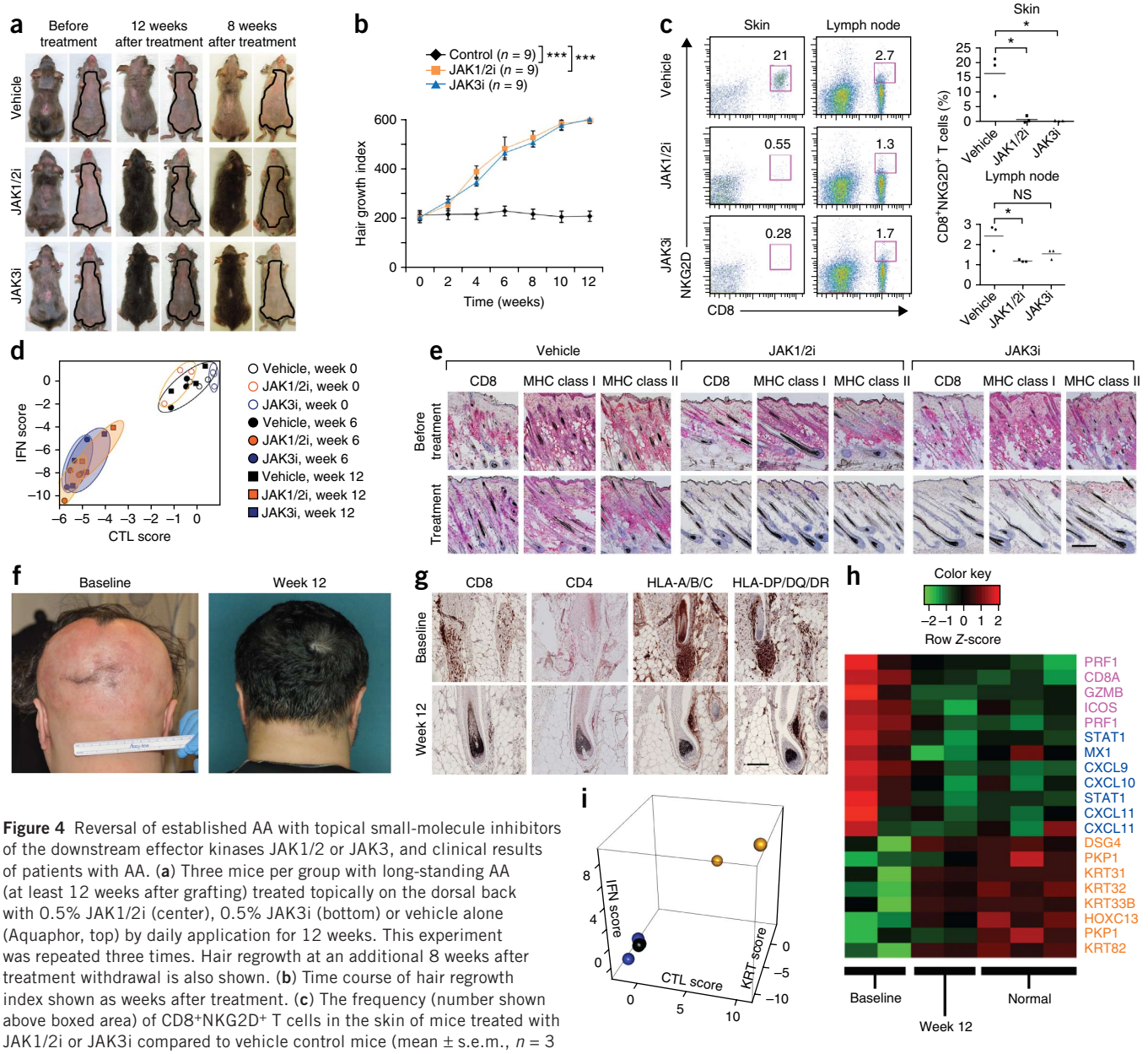


Figure 4 Reversal of established AA with topical small-molecule inhibitors of the downstream effector kinases JAK1/2 or JAK3, and clinical results of patients with AA. (**a**) Three mice per group with long-standing AA (at least 12 weeks after grafting) treated topically on the dorsal back with 0.5% JAK1/2i (center), 0.5% JAK3i (bottom) or vehicle alone (Aquaphor, top) by daily application for 12 weeks. This experiment was repeated three times. Hair regrowth at an additional 8 weeks after treatment withdrawal is also shown. (**b**) Time course of hair regrowth index shown as weeks after treatment. (**c**) The frequency (number shown above boxed area) of $CD8^+NKG2D^+$ T cells in the skin of mice treated with JAK1/2i or JAK3i compared to vehicle control mice (mean ± s.e.m., $n = 3$ per group, $^*P < 0.05$, $^{**}P < 0.01$, statistical methods described in the **Supplementary Methods**). NS, not significant. (**d**) The ALADIN score shows treatment-related loss of CTL and IFN signatures, given as log2 mean expression *Z*-scores as indicated in the **Supplementary Methods**. (**e**) Immunohistochemical staining of mouse skin biopsies shows treatment-related loss of expression of CD8 and MHC class I and II markers. Scale bar, 100 μm. (**f**) Treatment of patient 3 with AA, who had hair loss involving >80% of his scalp at baseline, with ruxolitinib and hair regrowth after 12 weeks of oral treatment. (**g**) Clinical correlative studies of biopsies obtained before treatment (baseline) and after 12 weeks of treatment of patient 2, including immunostains for CD4, CD8 and human leukocyte antigen (HLA) class I (A, B, C) and class II (DP, DQ, DR). Scale bar, 200 μm. (**h**,**i**) RNA microarray analysis from treated patients 1 and 2 with AA (before treatment versus after treatment versus 3 normal subjects) presented as a heatmap (**h**) and as a cumulative ALADIN index (**i**). KRT, hair follicle keratins.

To test whether inhibition of these signaling pathways would be therapeutically effective *in vivo*, we systemically administered ruxolitinib (**Fig. 3a–c**) and tofacitinib (**Fig. 3f–h**) at the time of grafting and found that they prevented the development of AA and the expansion of $CD8^+NKG2D^+$ T cells in all grafted recipients. The skin of mice treated with either drug showed no histological signs of inflammation (**Fig. 3d,i**). Global transcriptional analysis of whole-skin biopsies showed that both drugs also blocked the dermal inflammatory signature, as measured by Alopecia Areata Disease Activity Index (ALADIN, **Fig. 3e,j**), and Gene Expression Dynamic Index (GEDI) analysis (**Supplementary Fig. 10**)[23].

We next asked whether systemic tofacitinib treatment could reverse established disease by initiating therapy 7 weeks after grafting, a time point at which all mice had developed extensive AA. Systemic therapy resulted in substantial hair regrowth all over the body, reduced the frequency of $CD8^+NKG2D^+$ T cells and reversed histological markers of disease (**Supplementary Fig. 11**), all of which persisted 2–3 months after the cessation of treatment.

Next, to test a more clinically relevant route of delivery, we asked whether topical administration of protein tyrosine kinase inhibitors could reverse established AA in mice with kinetics similar to those of systemic delivery. In established disease, we found that topical ruxolitinib and topical tofacitinib were both highly effective in reversing disease in treated lesions (applied to back skin). A full coat of hair emerged in the ruxolitinib- or tofacitinib-treated mice by 7 weeks of treatment (data not shown), and we observed complete hair regrowth within 12 weeks following topical therapy (**Fig. 4a,b**). Topical therapy was associated with a markedly reduced proportion of $CD8^+NKG2D^+$ T cells in the treated skin and lymph node (**Fig. 4c**), normalization of the ALADIN transcriptional signature (**Fig. 4d**), reversal of histological markers of disease (**Fig. 4e**) and correction of the GEDI in all treated mice (**Supplementary Fig. 12**). Notably, untreated areas on the abdomen remained alopecic (**Fig. 4a** and **Supplementary Fig. 13**), demonstrating that topical therapy acted locally and that the observed therapeutic effects were not the result of systemic absorption. These effects were visible as early as 2–4 weeks after the onset of treatment (**Supplementary Fig. 14**) and persisted 2–3 months after the cessation of treatment (**Fig. 4a**).

To test the efficacy of JAK inhibitors in human subjects with AA, we treated three patients with moderate to severe disease orally with ruxolitinib, 20 mg twice daily. Ruxolitinib is currently FDA-approved for the treatment of myelofibrosis[24,25], a disease driven by wild-type and mutant JAK2 signaling downstream of hematopoietic growth factor receptors. In addition, small clinical studies using topical ruxolitinib in psoriasis[26] have demonstrated anti-inflammatory activity that may be due to interruption of the IL-17 signaling axis. All three ruxolitinib-treated patients exhibited near-complete hair regrowth within 3 to 5 months of oral treatment (**Fig. 4f** and **Supplementary Figs. 15** and **16**). Comparison of biopsies obtained at baseline and after 12 weeks of treatment demonstrated reduced perifollicular T cell infiltration, reduced follicular expression of human leukocyte antigen class I and class II expression (**Fig. 4g**) and normalization of the ALADIN inflammatory and hair keratin signatures following treatment (**Fig. 4h,i**).

Taken together, our data suggest $CD8^+NKG2D^+$ T cells promote AA pathogenesis, acting as cytolytic effectors responsible for autoimmune attack of the hair follicle (**Supplementary Fig. 17**). We postulate that IFN-γ produced by CD8 T cells leads to the collapse of immune privilege in the hair follicle[27], inducing further production of IL-15 (ref. 28) (**Supplementary Figs. 5** and **6**) and a feed-forward loop that promotes type I cellular autoimmunity. The clinical response of a small number of patients with AA to treatment with the JAK1/2 inhibitor ruxolitinib suggests future clinical evaluation of this compound or other JAK protein tyrosine kinase inhibitors currently in clinical development is warranted in AA[29].

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Microarray and RNA-seq data were deposited in Gene Expression Omnibus with accession numbers GSE45657, GSE45512, GSE45513, GSE45514, GSE45551 and GSE58573.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the National Alopecia Areata Registry, as well as M. Duvic, V. Price, M. Hordinsky and D. Norris, and the National Alopecia Areata Foundation. We thank J. Sundberg, T. Behrens, D. Bickers, J. O'Shea, T. Waldmann, B. Jabri, D. Raulet, L. Lanier, T. Spies, M. Hayden, R. Paus, P. Green, B. Lebwohl, D. Accili and C. Jahoda for stimulating discussions. We are grateful for clinical support from M. Furniss, C. Clark and G. Ulerio and expert technical assistance from M. Zhang, E. Chang, H. Lam and J. Huang. This work was supported in part by US Public Health Service National Institutes of Health NIAMS grants R01AR056016 (to A.M.C.) and R21AR061881 (to A.M.C and R.C.), a Shared Instrumentation Grant for the LSR II Flow Cytometer (S10RR027050) to R.C. and the Columbia University Skin Disease Research Center (P30AR044535), as well as the Locks of Love Foundation and the Alopecia Areata Initiative. J.E.C. is supported by the T32GM082271 Medical Genetics Training Grant (issued to A.M.C.). A.J., C.A.H., S.H. and A.d.J. are recipients of Career Development Awards from the Dermatology Foundation, and A.J. is also supported by the Louis V. Gerstner Jr Scholars Program.

AUTHOR CONTRIBUTIONS

L.X., Z.D. and A.J. were responsible in large part for performing the studies reported herein and participated in the design, execution and interpretation of the data. C.A.H. was responsible for establishing the C3H/HeJ graft model. A.d.J., S.H., G.M.D., L.R. and P.S. were involved in additional molecular and cell biological experiments. W.G. performed immunofluorescence and morphometric studies. L.P. and J.E.C. performed biostatistical analysis of all data sets. J.M.-W. was instrumental in human sample acquisition and analysis. A.M.C. and R.C. were responsible for conception, design, oversight, execution and interpretation of data for this study. All authors contributed to drafts, writing, figure preparation and editing of the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Gilhar, A., Etzioni, A. & Paus, R. Alopecia areata. *N. Engl. J. Med.* **366**, 1515–1525 (2012).
2. Petukhova, L. *et al.* Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. *Nature* **466**, 113–117 (2010).
3. Gilhar, A., Ullmann, Y., Berkutzki, T., Assy, B. & Kalish, R.S. Autoimmune hair loss (alopecia areata) transferred by T lymphocytes to human scalp explants on SCID mice. *J. Clin. Invest.* **101**, 62–67 (1998).
4. McElwee, K.J. *et al.* Transfer of $CD8^+$ cells induces localized hair loss whereas $CD4^+$/$CD25^-$ cells promote systemic alopecia areata and $CD4^+$/$CD25^+$cells blockade disease onset in the C3H/HeJ mouse model. *J. Invest. Dermatol.* **124**, 947–957 (2005).
5. Ito, T. *et al.* Maintenance of hair follicle immune privilege is linked to prevention of NK cell attack. *J. Invest. Dermatol.* **128**, 1196–1206 (2008).
6. Sundberg, J.P., Cordy, W.R. & King, L.E. Jr. Alopecia areata in aging C3H/HeJ mice. *J. Invest. Dermatol.* **102**, 847–856 (1994).
7. McElwee, K.J., Boggess, D., King, L.E. Jr. & Sundberg, J.P. Experimental induction of alopecia areata-like hair loss in C3H/HeJ mice using full-thickness skin grafts. *J. Invest. Dermatol.* **111**, 797–803 (1998).
8. Bertolini, M. *et al.* Abnormal interactions between perifollicular mast cells and $CD8^+$ T cells may contribute to the pathogenesis of alopecia areata. *PLoS ONE* **9**, e94260 (2014).

9. Best, J.A. *et al.* Transcriptional insights into the CD8$^+$ T cell response to infection and memory T cell formation. *Nat. Immunol.* **14**, 404–412 (2013).
10. Bezman, N.A. *et al.* Molecular definition of the identity and activation of natural killer cells. *Nat. Immunol.* **13**, 1000–1009 (2012).
11. Brajac, I., Gruber, F., Petrovecki, M. & Malnar-Dragojevic, D. Interleukin-2 receptor α-chain expression in patients with alopecia areata. *Acta Dermatovenerol. Croat. ADC* **12**, 154–156 (2004).
12. Fehniger, T.A. & Caligiuri, M.A. Interleukin 15: biology and relevance to human disease. *Blood* **97**, 14–32 (2001).
13. Ye, W., Young, J.D. & Liu, C.C. Interleukin-15 induces the expression of mRNAs of cytolytic mediators and augments cytotoxic activities in primary murine lymphocytes. *Cell. Immunol.* **174**, 54–62 (1996).
14. Meresse, B. *et al.* Reprogramming of CTLs into natural killer-like cells in celiac disease. *J. Exp. Med.* **203**, 1343–1355 (2006).
15. Saikali, P., Antel, J.P., Pittet, C.L., Newcombe, J. & Arbour, N. Contribution of astrocyte-derived IL-15 to CD8 T cell effector functions in multiple sclerosis. *J. Immunol.* **185**, 5693–5703 (2010).
16. Meresse, B. *et al.* Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. *Immunity* **21**, 357–366 (2004).
17. Freyschmidt-Paul, P. *et al.* Interferon-γ-deficient mice are resistant to the development of alopecia areata. *Br. J. Dermatol.* **155**, 515–521 (2006).
18. Gilhar, A., Kam, Y., Assy, B. & Kalish, R.S. Alopecia areata induced in C3H/HeJ mice by interferon-γ: evidence for loss of immune privilege. *J. Invest. Dermatol.* **124**, 288–289 (2005).
19. Freyschmidt-Paul, P. *et al.* Reduced expression of interleukin-2 decreases the frequency of alopecia areata onset in C3H/HeJ mice. *J. Invest. Dermatol.* **125**, 945–951 (2005).
20. O'Shea, J.J., Kontzias, A., Yamaoka, K., Tanaka, Y. & Laurence, A. Janus kinase inhibitors in autoimmune diseases. *Ann. Rheum. Dis.* **72** (suppl. 2), ii111–ii115 (2013).
21. Quintás-Cardama, A. *et al.* Preclinical characterization of the selective JAK1/2 inhibitor INCB018424: therapeutic implications for the treatment of myeloproliferative neoplasms. *Blood* **115**, 3109–3117 (2010).
22. Ghoreschi, K. *et al.* Modulation of innate and adaptive immune responses by tofacitinib (CP-690,550). *J. Immunol.* **186**, 4234–4243 (2011).
23. Eichler, G.S., Huang, S. & Ingber, D.E. Gene Expression Dynamics Inspector (GEDI): for integrative analysis of expression profiles. *Bioinformatics* **19**, 2321–2322 (2003).
24. Verstovsek, S. *et al.* Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N. Engl. J. Med.* **363**, 1117–1127 (2010).
25. Harrison, C. *et al.* JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N. Engl. J. Med.* **366**, 787–798 (2012).
26. Punwani, N. *et al.* Preliminary clinical activity of a topical JAK1/2 inhibitor in the treatment of psoriasis. *J. Am. Acad. Dermatol.* **67**, 658–664 (2012).
27. Paus, R., Nickoloff, B.J. & Ito, T.A. A 'hairy' privilege. *Trends Immunol.* **26**, 32–40 (2005).
28. Waldmann, T.A. The biology of interleukin-2 and interleukin-15: implications for cancer therapy and vaccine design. *Nat. Rev. Immunol.* **6**, 595–601 (2006).
29. Dolgin, E. Companies hope for kinase inhibitor JAKpot. *Nat. Rev. Drug Discov.* **10**, 717–718 (2011).

ONLINE METHODS

Mice. 7- to 10-week-old female C57BL/6 and C3H/HeJ mice (Jackson Laboratories, Bar Harbor, ME) were used and maintained under specific pathogen–free conditions. Experiments were performed in compliance with institutional guidelines as approved by the Institutional Animal Care and Use Committee of Columbia University Medical Center.

Transfer of alopecia areata using grafted C3H/HeJ skin. Normal-haired C3H/HeJ mice were grafted at 8 weeks of age (during the second telogen) with skin from a C3H/HeJ mouse that developed AA spontaneously, as described previously[7]. In brief, mice spontaneously affected with AA were euthanized, and full thickness skin grafts of approximately 2 cm in diameter were removed and grafted to normal-haired C3H/HeJ mice. All experiments included 10 grafted mice (5 treated and 5 untreated) and 3 sham grafted mice (mice grafted with autologous grafts). Hair loss typically began at around 4–6 weeks after grafting. No sham grafted mouse developed hair loss.

Flow cytometric analysis of skin and cutaneous lymph nodes. To make a single-cell suspension of mouse skin, fat was removed from the overlying skin in cold PBS and then incubated in collagenase type I (2 mg/ml in PBS) at 32 °C for 75 min. After digestion, the skin was minced in RPMI/10% FBS, filtered through a 70-μM cell strainer and centrifuged at 1100*g* for 5 min. The pellet was resuspended in RPMI/10% FBS, filtered through a 40-μM cell strainer and spun at 400*g* for 5 min. The pellet was resuspended in FACS buffer (PBS/5% BSA), DAPI to gate on live cells and staining antibodies (listed in **Supplementary Methods**). Cutaneous lymph nodes were pooled, minced in RPMI, filtered through a 40-μM cell strainer, centrifuged at 400*g* for 5 min, stained and analyzed on a BD LSR II flow cytometer.

Transfer of T cell populations into recipient C3H/HeJ mice. For positive selection of T cell populations, lymph node cells were obtained from 5 C3H/HeJ alopecic mice, stained with anti-CD4, anti-CD8 and anti-NKG2D antibodies) and then sorted (BD Influx) to obtain two fractions: $CD8^+NKG2D^+$ T cells and $CD8^+NKG2D^-$ T cells. Antibody dilutions are given in the **Supplementary Methods**. Three to five 7-week-old C3H/HeJ mice per group were injected subcutaneously with two million sorted cells of each population. For negatively selected populations, $NKG2D^+$ cells were depleted by incubating total lymph node cells from 3 alopecic C3H/HeJ mice with biotinylated anti-NKG2D (CX5) and then with streptavidin-conjugated beads (Miltenyi) before removal on a Miltenyi magnetic column. Five million cells (either CD8/NKG2D-depleted or total lymph node cells) were suspended in 100 ul PBS and transferred into each of 5 mice by subcutaneous injection.

Prevention and treatment studies in mice. For prevention studies, mice were treated beginning on the day of grafting (n = 5–10 mice per group). For anti-IFN-γ experiments, i.p. injections of hamster isotype control IgG or hamster polyclonal anti-IFN-γ IgG (BioXCell) 300 μg in 100μl PBS were administered ten times weekly for 12 weeks. For anti–IL-2 experiments, i.p. injections of control rat isotype control IgG or simultaneous administration of two anti-IL-2 rat monoclonal antibodies (BioXCell) (250 μg clone S4B6 and 250 μg JES6-1 together in 100 μl PBS) were administered three times weekly for 12 weeks. For anti–IL-15-Rβ experiments, i.p. injections of rat isotype control IgG or anti-IL-15Rβ antibody (Biolegend, clone TM-β1) (200 μg in 100 μl PBS) were administered two times weekly for 12 weeks. For JAK1/2i experiments, mice were administered vehicle (0.5% methylcellulose; Sigma-Aldrich) or vehicle-containing 50 mg/kg of JAK1/2i ruxolitinib (ChemieTek) daily by oral gavage for 12 weeks. For JAK3i experiments, mice were implanted subcutaneously with Alzet osmotic mini-pumps (pumps, model 2004, Durect Corporation) on the back of each mouse to deliver vehicle (poly(ethylene glycol) (PEG)300) or vehicle containing the JAK3i tofacitinib (Abmole) at 15 mg/kg/day for 12 weeks.

For topical treatment studies, grafted mice with long-standing AA (more than 8 weeks) were treated once daily for 12 weeks to affected skin on the dorsal back with vehicle (10% DMSO in Aquaphor) or vehicle containing JAK inhibitors, initially dissolved in DMSO and then diluted 1:10 in Aquaphor, to achieve 0.5% JAKi ointment). Full-thickness skin biopsies were excised from the dorsal surface of each mouse at interim time points, and skin samples were either snap frozen in liquid nitrogen for RNA extraction or snap frozen in OCT for immunostaining. Hair status was examined twice weekly.

Human clinical studies. We initiated a single center, proof-of-concept clinical trial in the Clinical Trials Unit in the Department of Dermatology at the Columbia University Medical Center entitled "An Open-Label Pilot Study to Evaluate the Efficacy of ruxolitinib in Moderate to Severe Alopecia Areata," ClinicalTrials.gov identifier NCT01950780. The protocol for this intervention trial was reviewed and approved by the Institutional Review Board at Columbia University and conducted under the Declaration of Helsinki principles. Informed consent was received before inclusion in the study. Eligibility criteria included >30% hair loss for at least 3 months in duration with no evidence for hair regrowth at the time of enrollment. The first three treated patients are described here. The ruxolitinib dose was 20 mg orally twice daily for 3–6 months. Skin punch biopsies were obtained at baseline and after 12 weeks of treatment. Consent for photography was obtained for the patients shown in **Figure 4** and in the **Supplementary Figures 15** and **16**.

Immunohistochemistry and immunofluorescence. 8 μM acetone-fixed frozen mouse skin or human skin sections were air-dried and stained overnight with anti-mouse or anti-human antibodies (see **Supplementary Methods**) at 4 °C in a moist chamber. Human hair follicles were microdissected and embedded in OCT compound before sectioning and staining (see **Supplementary Methods**).

Primary dermal sheath and lymphokine-activated killer (LAK) cell culture. Dermal sheath (DS) cells were isolated from microdissected mouse vibrissa follicles and cultured in 20% FBS DMEM with 5 ng/ml murine FGF (Pepro Tech). LAK cells were generated from bulk splenocytes plated at 4×10^6 in 6-well plates with 50 ng/ml murine IL-15 (Pepro Tech), 50 nM JAK3i (tofacitinib) or 50 ng/ml murine IL-15 plus 50 nM JAK3i and incubated at 37 °C in a 5% CO_2 incubator for 96 h.

***In vitro* cytotoxicity assays.** Determination of specific killing of target cells was performed using CFSE-labeled DS cells as targets mixed with different ratios of effector cells incubated for 5 h at 37 °C 5% CO_2 with or without neutralizing rat anti-mouse NKG2D antibody (20 ug/ml) (Biolegend, CX5). Specific lysis of DS cells was determined flow cytometrically by measuring cell death of CFSE + DS cells using Annexin V/7-AAD.

Gene expression sample preparation in human and mouse skin and T cells. Total RNA was isolated using the miRNeasy Mini Kit (Qiagen Inc., Valencia, CA, USA) with on-column DNA digestion using the RNase-free DNase set (Qiagen, Inc.). For RNA-seq analysis $CD3^+CD8^+CD44^+NKG2D^+$ and $CD3^+CD8^+CD44^+NKG2D^-$ cells were flow-sorted from lymph nodes of alopecic C3H/HeJ mice. RNA was extracted as above and prepared for RNA-seq using the TruSeq RNA Sample Prep Kit v2. Samples were sequenced on the HiSeq 2000 sequencer (Illumina, San Diego, CA) for 50 cycles. RNA-seq files were demultiplexed by the Rockefeller University Genomics Core Facility.

For global transcriptional profiling in mouse skin, total extracted RNA was processed using the 3′ IVT Express Kit from Affymetrix. Resulting biotinylated cDNA samples were hybridized to the Mouse Genome 430 2.0 gene chips and subsequently washed, stained with streptavidin-phycoerythrin and scanned on an HP GeneArray Scanner (Hewlett-Packard Company, Palo Alto, CA). For gene expression studies, mice grafted with autologous healthy skin were included as sham-operated controls.

For human AA samples, perilesional punch biopsies from 5 patients with patchy alopecia areata who were not undergoing local or systemic treatments were collected and compared to scalp biopsies from 5 unrelated unaffected individuals. All procedures were performed under Institutional Review Board–approved protocols at Columbia University and conducted under the Declaration of Helsinki principles. Informed consent was received before inclusion in the study. Extracted total RNA was reverse transcribed and amplified using the Ovation RNA Amplification V2 kit (NuGEN Technologies, Inc.,

San Carlos, CA). Amplified cDNA was biotinylated with the Encore Biotin Module (NuGEN Technologies) and then hybridized to the U133 Plus 2.0 gene chips. RT-PCR confirmations done as described in the **Supplementary Methods**. Primer sequences are given in **Supplementary Tables 4** and **5**.

Statistical analyses. Statistical methods for each figure are given in the **Supplementary Methods**. No statistical method was used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment. The experiments were not randomized.

Supplementary Note for:

Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition

Luzhou Xing[1*], Zhenpeng Dai[2*], Ali Jabbari[2*], Jane E. Cerise[2,3], Claire A. Higgins[2], Weijuan Gong[2], Annemieke de Jong[2], Sivan Harel[2], Gina M. DeStefano[2,4], Lisa Rothman[2], Pallavi Singh[2], Lynn Petukhova[2], Julian Mackay-Wiggan[2], Angela M. Christiano[2,5*] and Raphael Clynes[1,2,6*]

[1]Departments of Pathology [2]Dermatology, [3]Psychiatry, [4]Epidemiology, [5]Genetics & Development and [6]Medicine, Columbia University, New York, NY.

*These authors contributed equally to this study.

Nature Medicine: doi:10.1038/nm.3645

METHODS

Mice

C3H/HeJ mouse strain (Jackson Laboratories, Bar Harbor, ME) was used for all animal studies. Only female mice were used. Mouse recipients of alopecic skin grafts were aged 7-10 weeks at the time of grafting. For prevention experiments, drug administration began the day after grafting. For systemic treatment studies, drug administration was initiated approximately 3 months after mice lost their hair. For topical treatment studies, drug administration was initiated 20 weeks following grafting. All animal procedures were done according to protocols approved by the Columbia University Medical Center Institutional Animal Care and Use Committee.

Human studies

All human studies have been approved by the Columbia University Medical Center Institutional Review Board and were conducted under the Declaration of Helsinki principles. Informed written consent was received from participants before inclusion in the study.

Clinical Evaluation of Oral Ruxolitinib in Alopecia Areata

We initiated a single center, proof-of-concept clinical trial in the Clinical Trials Unit in the Department of Dermatology at the Columbia University Medical Center entitled “An Open-Label Pilot Study to Evaluate the Efficacy of RUXOLITINIB in Moderate to Severe Alopecia Areata” (clinicaltrials.gov identifier: NCT01950780).

Nature Medicine: doi:10.1038/nm.3645

The primary efficacy endpoint of this initial pilot study is the proportion of responders achieving 50% or greater regrowth at the end of treatment compared to baseline. Secondary endpoints include the changes in hair growth both during and after treatment measured as a continuous variable; patient global assessments; quality of life assessments; and durability of response following treatment cessation.

Inclusion criteria included 30 to 95% hair loss due to alopecia areata (AA) as measured by SALT score[30]; hair loss duration of at least 3 months; stable hair loss without active evidence of regrowth; subject age 18-75 years.

Exclusion criteria included active scalp disease other than AA; medical history that might increase the risks related to ruxolitinib e.g. hematologic, infectious, immune related diseases or malignancies; current treatment with any modality that might affect AA response; medications known to interact with ruxolitinib; pregnancy; etc.

Subjects on study are treated with oral ruxolitinib 20mg BID for at least 3 months. The patients in this manuscript have achieved over 90% regrowth. Skin punch biopsies (4mm) were obtained at baseline and after 12 weeks of treatment.

Antibodies used for mice treatment, flow cytometry, immunostaining and western blot analysis

All antibodies used in these studies are listed in table form below.

Flow cytometric analysis used the following anti-mouse antibodies: CD3 (17A2, Ebioscience), CD4 (GK1.5, BD), CD8α (53-6.7, BD), CD8β (YTS156.7.7, Biolegend), NKG2D (CX5, Ebioscience), NKG2A/C/E (clone 20d5, Ebioscience), CD44 (IM7, BD), CD45 (30-F11, BD), CD49b (Dx5, BD), CD62L (MEL-14, BD), CD69 (H1.2F3, BD),

CD103 (2E7, eBioscience), IFNγ (XMG1.2, Ebioscience), Granzyme B (NGZB, eBioscience), Rae-1 (186107, R&D).

For immunohistochemical studies of mouse skin, 8 μM methanol-fixed frozen skin sections were stained with primary rat antibodies (Biolegend) including: anti-CD8 (clone 53-6.7), Biotin anti-MHC class I (clone 36-7.5), anti-MHC class II (clone M5/114.15.2). Biotinylated goat anti-rat IgG (Life Technologies) was used as secondary antibody. For immunofluorescence studies anti-H60 (R&D, clone 205326), anti-Pan Rae-1(R&D, clone 186107), anti-NKG2D (R&D clone 191004), anti-IL-15 (SCBT, H-114), anti-IL-15 RA (SCBT, N-19), anti-K71 (Abcam), primary antibody were used in immunofluorescence. Alexa Fluor 488 or Alexa Fluor 594-conjugated goat anti-Rat, donkey anti-Rabbit or donkey anti-Goat antibody was used as secondary antibody (Life Technologies).

For immunohistochemical studies of human skin, 5μM formalin fixed and paraffin skin section were used. After heat antigen retrieval, skin sections were stained with primary anti-human antibodies including: anti-CD8(Abcam ab4055), anti-CD4(,Leica clone 1-F6), HLA Class 1 ABC(Abcam clone EMR8-5), HLA-DR/DP/DQ(SCBT clone CR3/43). ImmPRESS HRP Anti-Rabbit Ig or Mouse Ig (Peroxidase) Polymer (Vector Lab) were used as secondary antibody.

Human hair follicles were microdissected and embedded in OCT compound prior to sectioning and staining. 8 μM methanol-fixed frozen sections were stained with anti-IL-15 (SCBT, H-114) and anti-IL-15 RA (SCBT, N-19) or anti-IL-15 RB (SCBT, C-20) and CD8 (SCBT, C8/144B) followed by staining with Alexa Fluor 488 or Alexa Fluor

594-conjugated secondary antibody (Life Technologies). All images were captured with an SDRC Zeiss Exciter Confocal Microscope.

For western blotting, samples with treatment were resolved by 4-12% SDS-PAGE (Life Technologies) and then transferred to Westran PVDF membranes (GE Healthcare life Sciences). Blots were probed with the following Abs (All from Cell Signaling Technology): anti-phospho-STAT1 (Tyr701), anti-phospho-STAT5 (Tyr694), anti-STAT1 and anti-STAT5.

Antibodies for in vivo *treatment*

Mouse antibody	**Company**	**Clone**	**Cat No.**
Anti-IL15 Rβ	Biolegend	TM-β1	123204
IL-2	BioXcel	S4B6-1	BE0043-1
IL-2	BioXcel	JES6-1A12	BE0043
IFN-γ	BioXcel	H22	BE0254
IL-21	Ebioscience	FFA21	16-7211-85

Antibodies for flow cytometry (1/100 dilution)

Mouse antibody	**Company**	**Clone**	**Cat No.**
CD3	Ebioscience	17A2	17-0032
CD4	BD	GK1.5	560181
CD8a	BD	53-6.7	560469
CD8b	Biolegend	YTS156.7.7	126610
NKG2D	Ebioscience	CX5	12-5882
NKG2A/C/E	Ebioscience	20d5	13-5896
CD44	BD	IM7	553133
CD45	BD	30-F11	552848
D49b	BD	Dx5	553857
Cd62L	BD	MEL-14	553152
CD69	BD	H1.2F3	557392
CD103	Ebioscience	2E7	17-1031
IFNγ	Ebioscience	XMG1.2	11-7311

Nature Medicine: doi:10.1038/nm.3645

Pan Rae-1	R&D systems	186107	MAB17582
Granzyme B	Ebioscience	NGZB	11-8898

Antibodies for immunostaining and western blot (1/100 dilution unless otherwise noted)

Human antibody	**Company**	**Clone**	**Cat No.**
CD3	Abcam	PS1	Ab699
CD8	SCBT	C8/144B	Sc-53212
CD4	Leica	1-F6	CD4-1F6-L-CE
HLA Class I ABC	Abcam	EMR8-5	ab70328
HLA-DR/DP/DQ	SCBT	CR3/43	sc-53302

Mouse antibody	**Company**	**Clone**	**Cat No.**
CD8	Biolegend	53-6.7	100702
MHC-class I	Biolegend	36-7.5	114903
MHC-class II	Biolegend	M5/114.15.2	107602
H60	R&D systems	205326	MAB1155
Pan Rae-1	R&D systems	186107	MAB17582
NKG2D	R&D systems	191004	MAB1547

Mouse/Human antibody	**Company**	**Clone**	**Cat No.**	**IF/IHC Dilution**
IL-15	SCBT	polyclonal H-114	sc-7889	
IL-15RA	SCBT	polyclonal N-19	sc-1524	
Phospho-Stat1 (Tyr701)	Cell signaling	D4A7	7649	
Phospho-Stat3 (Tyr705)	Cell signaling	D3A7	9145	1/200
Phospho-Stat5 (Tyr694)	Cell signaling	C11C5	9359	1/400
Stat1	Cell signaling	polyclonal	9172	
Stat5	Cell signaling	polyclonal	9363	
K71	Abcam	polyclonal	Ab133817	

STAT1, STAT5, pSTAT1 and pSTAT5 ab's were diluted 1/1000 for western blots.

IL-15 and IL-15RA staining blocking reagents

Blocking reagent	**Company**	**Cat No.**
IL-15	Peprotech	AF-200-15
IL-15 RA blocking peptide	SCBT	sc-1524 P

Nature Medicine: doi:10.1038/nm.3645

RNA-Seq analysis

Samples were sequenced on the HiSeq 2000 sequencer (Illumina, San Diego, CA) for 50 cycles. RNA-Seq files were demultiplexed by the Rockefeller University Genomics Core Facility. Quality control of the sample fastq files was performed using fastqc[31]. TopHat[32] was used to map transcripts to the UCSC mm9 reference genome from iGenome. The RefSeq gene annotation packaged with this iGenome version of the UCSC mm9 were used. The htseq-count utility from the HTSeq package was used to convert TopHat bam files to counts that could be used as input for downstream analysis of differential expression with edgeR[33]. Absent genes were removed and a pseudocount of 1 was added in order to avoid division by zero in downstream analysis. EdgeR was used to identify differentially expressed genes using a matched pairs design with three biological replicates.

Microarray Analysis

Quality Control, Preprocessing

For the mouse cDNA samples were hybridized to the Mouse Genome 430 2.0 gene chips and subsequently washed, stained with streptavidin-phycoerythrin, and scanned on an HP GeneArray Scanner (Hewlett-Packard Company, Palo Alto, CA). For the human, amplified cDNA was hybridized to the Human Genome U133 Plus 2.0 gene chips.

Microarray quality control and preprocessing were performed using BioConductor in R. Preprocessing of the three experiments, 1) spontaneous AA mice vs. normal

mice, 2) prevention mice with three treatments vs. placebo and sham-operated mice, and 3) treatment mice for two treatments vs. placebo were performed separately using the same pipeline.

Quality control was performed using the affyanalysisQC package from http://arrayanalysis.org/. AffyanalysisQC uses the R/BioConductor packages: affy, affycomp, affypdnn, affyPLM, affyQCReport, ArrayTools, bioDistm biomaRt, simpleaffy, and yaqcaffy to perform QC within a single script. RMA normalization[34] was performed on each experimental group separately. Batch effect correction using ComBat was required for the prevention experiments. Batches, treatments and time points were modeled treating each treatment group effect as constant over time, and grouping the PBS controls in groups reflecting both treatment and time.

In addition to the preprocessing that was done for the mouse skin samples, Harshlight was used to correct for image defects for the human skin samples.

Data Deposition

Microarray and RNA-seq data was deposited in Gene Expression Omnibus, accession numbers GSE45657, GSE45512, GSE45513, GSE45514, GSE45551, and GSE58573.

Identification of Gene Signatures

Differential expression analysis

Initial analysis of differential gene expression was performed on the spontaneous mouse 3x3 and the human 5x5 data sets using limma[35]. A threshold of 1.5 fold change and unadjusted p-value of 0.05.

Unsupervised analysis

Hierarchical clustering was performed using Cluster[36] on the 363 genes from the human 5x5 and 583 genes from the spontaneous mouse 3x3 that met the threshold abs(logFC) > 1, unadjusted p-value <= 0.05. Genes were median centered and normalized. Spearman rank correlation was used as the similarity measure and average linkage was used to perform row (genes) and column (sample) clustering. Visualization of the hierarchical clusters was performed with java TreeView[37]. Gene Expression Dynamic Index (GEDI) analysis was used to visualize how "metagenes" identified with a self organizing map algorithm vary across samples[38]. Metagenes are clusters of genes that show similar expression patterns across samples and that are assigned to a single pixel in a two dimensional grid. Neighboring pixels demonstrate similar expression patterns to one another.

RT-PCR Validation

Predicted differentially expressed genes in human and mouse were confirmed using RT-PCR. First-strand cDNA was synthesized using a ratio of 2:1 random primers: Oligo (dT) primer and SuperScript III RT (Invitrogen) according to the manufacturer's instructions. qRT-PCR was performed on an ABI 7300 machine and analyzed with ABI Relative Quantification Study software (Applied Biosystems, Foster City, CA, USA). Primers were designed according to ABI guidelines and all reactions were performed

using *Power* SYBR Green PCR Master Mix (Applied Biosystems), 250 nM primers (Invitrogen) and 20 ng cDNA in a 20μL reaction volume. Primer sequences are provided in Supplementary Tables 4 and 5. The following PCR protocol was used: step 1: 50°C for 2 min; step 2: 95°C for 10 min; step 3: 95°C for 15 s; step 4: 60°C for 1 min; repeat steps 3 and 4 for 40 cycles. All samples were run in quadruplicate for three independent runs and normalized against an endogenous internal control as indicated.

ALADIN scores

The IFN and CTL signatures were used to develop a bivariate score statistic. Individual signature IFN and CTL scores were determined following procedures used in human SLE[39,40]. The sets of genes selected to comprise our IFN and CTL signatures were CD8A, GZMB, and ICOS for the CTL signature, and CXCL9, CXCL10, CXCL11, STAT1, and MX1 for the IFN signature. The scores for the prevention mice were calculated in relation to the sham mice; whereas, the scores for the topical treatment experiments were calculated relative to all the samples at week zero. Based on our human studies (unpublished), ALADIN was further extended to include a hair keratin (KER) signature. The set of genes selected to comprise the KER signature are DSG4, HOXC31, KRT31, KRT32, KT33B, KRT82, PKP1, and PKP2. The ALADIN scores for the baseline and 12 week skin biopsies obtained from subjects enrolled in the oral Ruxolitinib clinical trial were calculated relative to the healthy controls at baseline.

Power analysis

For the analysis of response to treatment, we performed a two-sample comparison of proportions power calculation for group sample sizes of five each for

treated and placebo mice for the case when the true proportion in population 1 (the treatment group) expected to respond to treatment is 0.95 and the true proportion in population 2 (the placebo group) expected to respond is 0.20. At a significance level of alpha = 0.05, using Barnard's exact test we calculated a power of 0.803 for a one-sided test to detect a difference of proportions when the proportions for the two populations are 0.95 and 0.20 with group sample sizes equal to five each. In some cases in which fewer than 5 animals per group were present per experiment, multiple experiments were collapsed in order to ensure statistical power.

Statistical Analysis of Treatment Effects

Mice were expected to exhibit alopecia 4-12 weeks after grafting of alopecic skin. Experiments in which control mice failed to demonstrate hair loss by 8 weeks were aborted. For the prevention experiments, a time-to-event survival analysis for interval censored data was performed. The survival and interval packages in R were used to perform log-rank tests. Hair growth index was calculated as described[11].

For the treatment experiments (Figure 4b), the R package nparLD was used to test the hypothesis that there exists a treatment by time interaction. Analyses were performed using the hair growth index from three replicate experiments containing three mice from each treatment and placebo group for a total of nine mice from each group. A F1-LD-F1 design was employed. For the JAK1/2i treatment vs. placebo, the hypothesis of no interaction, i.e., parallel time profiles, is rejected at the 5% level using both the Wald-Type Statistic and the ANOVA-Type Statistic with the p-values of 4.40e-21 and 3.35e-18, respectively. For the JAK3i Treatment vs Placebo, the hypothesis of no interaction, i.e., parallel time profiles, is rejected at the 5% level using both the Wald-

Nature Medicine: doi:10.1038/nm.3645

Type Statistic and the ANOVA-Type Statistic with the p-values of 1.45e-30 and 2.42e-21, respectively.

All mice were included in survival (time-to-event) analysis statistics. For lymph node and skin cell analysis, biopsy was harvested at the indicated time points following treatment in parallel with control mice. In the IFN-γ- and IL-2- neutralization experiments one out of five control mice that did not exhibit hair loss was not included in the photographs (Figures 2b, 2e). These mice were not sacrificed in order to continue to monitor for hair loss, but for statistical purposes for skin cell analysis (Figure 2b, 2e), these unanalyzed samples were assigned a cell count value of 0% $CD8^+NKG2D^+$ cells to allow for a rigorous and conservative statistical comparison with treated mice.

No randomization was used and the investigators were not blinded to the group allocation during the experiments or when assessing the outcomes.

Unpaired parametric two-sided t-tests were used to test for differences in means and frequencies between treated and untreated groups. For statistical purposes, we assume all variances to be the same for each group.

Interval censored log-rank tests were used to perform all time to event survival analysis. This test properly accounts for data where the exact event time is not known but the event is known to fall within some interval.

Nonparametric longitudinal data analysis was used to test for response x time interactions. These methods are particularly suited for small sample size.

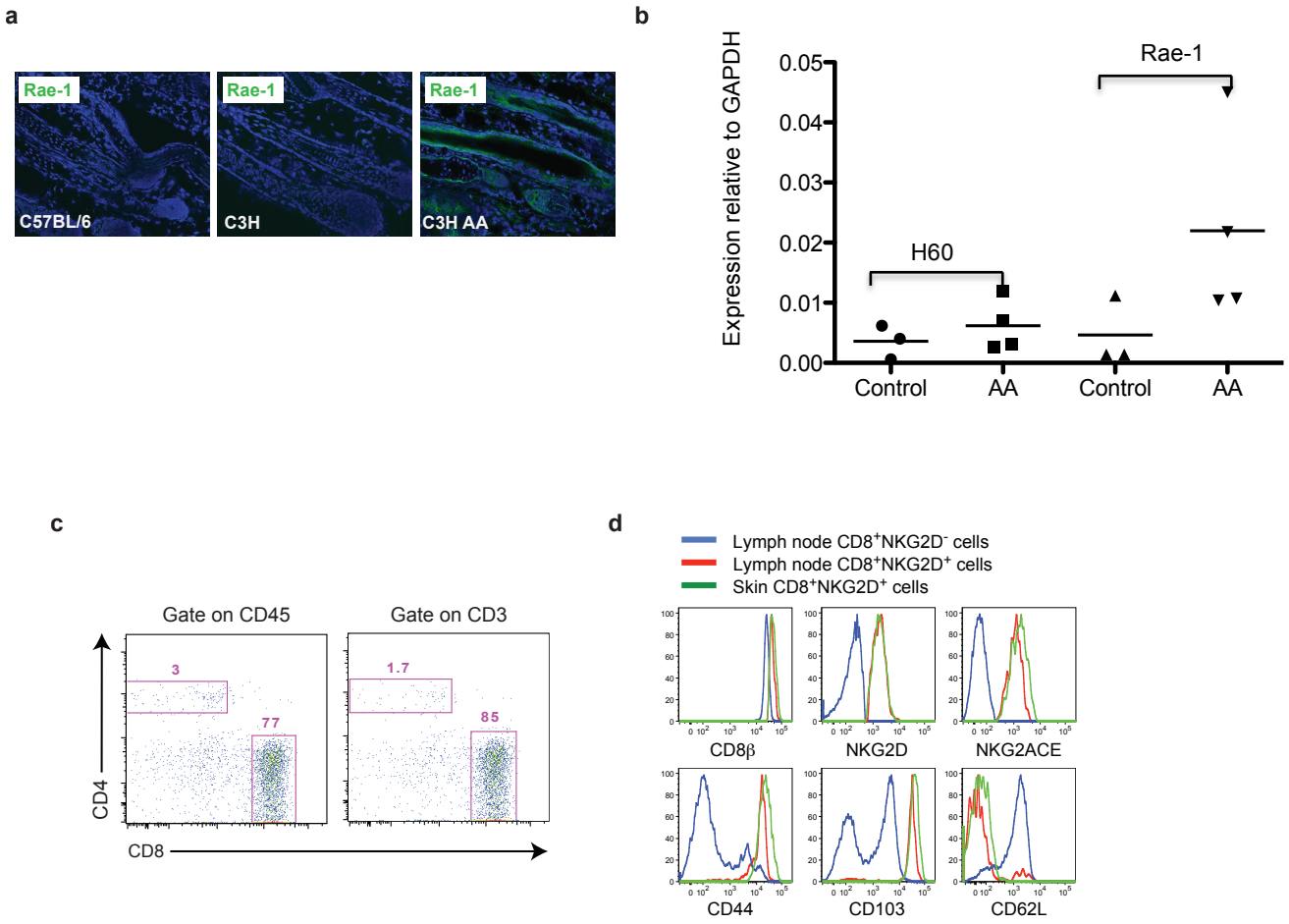
Sample sizes, number of replicates, and statistical tests among experiments

Figure	n_control	n_exp	experiment	statistic	p-value
1c	3+3+3+3	6+6+6+6	C3HAA/C3H cell counts	t	< 0.0001
1d	2+2+2	5+5+4	C3HAA/C3H skin	t	< 0.0001
1d	3+3+2	4+4+4	C3HAA/C3H lymph node	t	< 0.0001
1f	3	3	primary cell culture	N/A	NA
2b	5	5	α-IFNg, mice with hair loss	log-rank	0.047
2b	5	5	α-IFNg Skin cell counts	t	0.0228
2e	5	5	α-IL2, mice with hair loss	log-rank	0.048
2e	5	5	α-IL2 Skin cell counts	t	0.0091
2h	5+4+3	5+4+3	α-IL15Rb, mice with hair loss	log-rank	1.11E-05
2h	2+2+1	2+2+2	α-IL15Rb Skin cell counts	t	< 0.0001
3b	4+6	4+6	JAK1/2i, mice with hair loss	log-rank	0.00041
3c	2+3	3+3	JAK1/2i Skin cell counts	t	0.0003
3c	2+3	4+5	JAK1/2i lymph node cell counts	t	< 0.0001
3g	7	5	JAK3i, mice with hair loss	log-rank	0.0025
3h	2+2	2+3	JAK3i Skin cell counts	t	0.0002
3h	2+2	2+3	JAK3i lymph node cell counts	t	0.0049
4b	3+3+3	3+3+3	JAK1/2i / JAK3i / Vehicle	nonparLD	Jak1/2i 4.4e-21, Jak3i 1.5e-30
4c	3	3	JAK1/2i / JAK3i / Vehicle Skin	t	Jak1/2i 0.017, Jak3 0.015
4c	3	3	JAK1/2i / JAK3i / Vehicle lymph node	t	Jak1/2i p=0.0297, Jak3i p=0.0908

For *in vivo* studies data are provided as cumulative data. The number of replicates are provided as shown above; For example “3+3+3+3” refers to four separate experiments each including three experimental mice. For *in vitro* studies (Figures 1f, 3d, and 3e), experiments were performed in triplicate.

Nature Medicine: doi:10.1038/nm.3645

Supplementary Figures

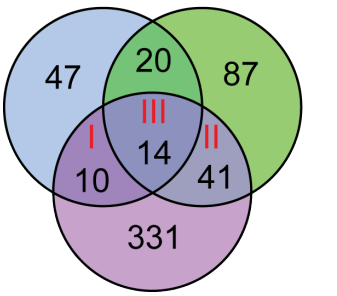


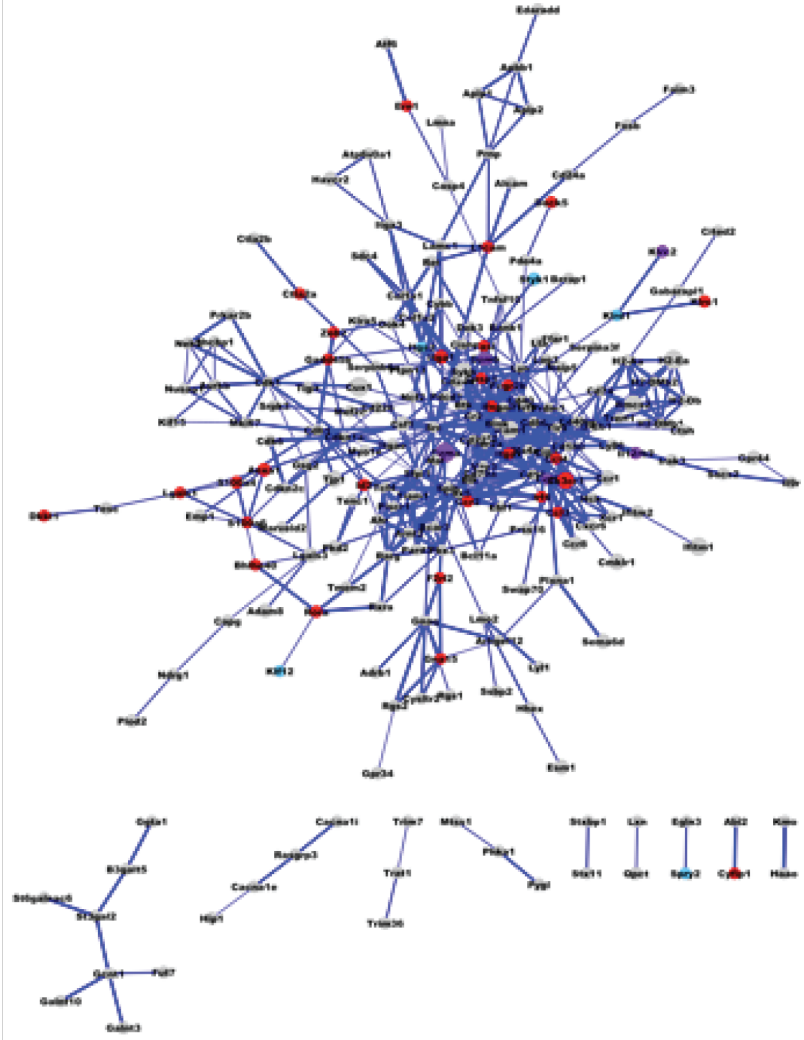
Supplementary Figure 1. NKG2D-NKG2DL expression in mouse AA

a, Rae-1 is upregulated in the AA HF. Immunostaining of lesional using a pan-Rae-1 antibody in C3H/HeJ alopecic mice, unaffected C3H/HeJ mice and C57Bl.6 mice. **b,** H60 and Rae-1 are overexpressed in AA. NKG2DL expression in alopecic lesional skin from C3H/HeJ mice compared with non-lesional skin from unaffected C3H/HeJ mice. RT-PCR data from cDNA from 3 mice are shown and represented as relative to GAPDH**. c,** CD4 T cells are infrequent in AA lesional skin: Flow cytometric evaluation of lesional alopecic skin. Quantitation of CD4 and CD8 T cells as a percentage of total gated $CD3^+$ or $CD45^+$ cells. **d,** Effector memory immunophenotype of $CD8^+NKG2D^+$ T cells are similar in lesional skin and in the cutaneous draining lymph node. Gated $CD8alpha^+NKG2D^+$ T cells are displayed for CD8beta, NKG2D, NKG2A/C/E, CD44, CD103 and CD62L expression.

Nature Medicine: doi:10.1038/nm.3645

Venn Diagram Group	Gene overlap
Group I (NKG2D$^+$CD8$^+$ T cell and NK cell (Bezman))	Arsb, Gem, Itga2, Klf12, Klrd1, Spry2, Styk1, Sytl3, 4831426I19Rik, A930038C07Rik
Group II (CD8$^+$NKG2D$^+$ T cell and CD8$^+$ cell (Best))	Anxa1, Atp2b4, Bhlhe40, Casp1, Ccl3, Ccl4, Ccr2, Crybg3, Ctla2a, Cx3cr1, Cyfip1, Dennd5a, Dkkl1, Dock5, Ern1, F2rl2, Fam129a, Fasl, Fcgr2b, Gadd45b, Gna15, Gzmk, Id2, Ifng, Irf4, Itga1, Itgax, Kcnk5 Klrc1, L1cam, Lgals1, Ncald, Ptprj, Rnf216, Rora, S100a4, S100a6 Soat2, Ttc39c, Zdhhc2, Zeb2
Group III (CD8$^+$NKG2D$^+$ T cell and CD8$^+$ cell (Best) and NK cell (Bezman))	Car5b, Ccl5, Fbxl2, Gm11435, Gzma, Gzmb, Il12rb2, Klrc2, Klrc3, Klre1, Klrk1, Osbpl3, S1pr5, Slc25a24



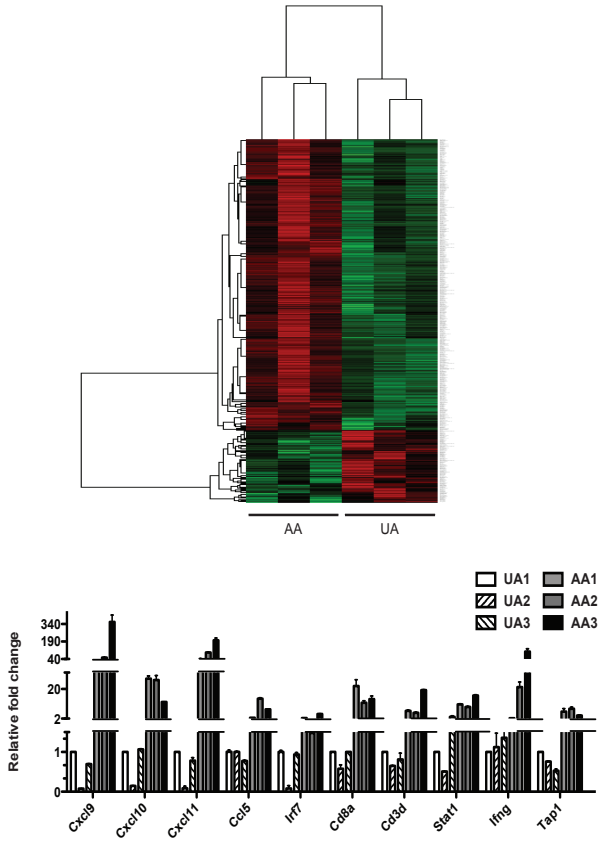


Supplementary Figure 2. Transcriptional profile of CD8$^+$NKG2D$^+$ cells in mouse AA.

Nature Medicine: doi:10.1038/nm.3645

Top and bottom left panels, List of genes and Venn diagram showing overlap of our CTL gene expression with those in the literature[41,42] Bottom right panel, Network Map of differentially expressed upregulated genes in $CD3^+$ $CD8^+NKG2D^+$ LN cells vs. $CD8^+NKG2D^-$ cells. String.db was used to create a biological interaction score matrix with the differentially expressed genes. The network map was created using cytoscape; only biological interactions >0.75 were used. Nodes represent genes, and edges represent biological interactions as derived from string.db. Node size is proportional to fold change, and edge width is proportional to biological interaction.

Signature	Human	Mouse
CTL	GZMA(3.65x), CD8A(3.04x), GZMB(2.75x), GZMK(2.68x), PRF1(2.30x), LCP2(2.29x), CD2(2.22x), IL7(2.20x), THEMIS(2.14x)	Gzma(56.37x), Cd8a(9.05x), Gzmb(21.74x), Gzmk(2.61x), Prf1(1.62x), Lcp2(2.32x), Cd2(3.32x), Il7(1.99x), Themis(1.84x)
IFN	CXCL10(12.37x), CXCL9(9.53x), MMP12(4.04x), IFI44(3.03x), SPP1(2.69x), IRF8(2.64x), PTPRC(2.63x), CCL2(2.54x), RSAD2(2.49x), CCL5(2.36x), IFIT2(2.28x), C1S(2.28x), TLR3(2.25x), IFIT3(2.19x), OAS2(2.12x), GBP1(2.11x), XCL1(2.05x), CCR2(2.04x), CXCL11(1.86x), IFNG(1.57x), STAT1(1.53x), JAK1(0.85x)	Cxcl10(37.58x), Cxcl9(42.30x), Mmp12(10.18x), Ifi44(20.58x), Spp1(10.61x), Irf8(4.44x), Ptprc(2.11x), Ccl2(6.12x), Rsad2(3.20x), Ccl5(27.95x), Ifit2(2.39x), C1s(2.49x), Tlr3(2.34x), Ifit3(4.75x), Oas2(3.97x), Gbp1(17.81x), Xcl1(2.88x), Ccr2(4.66x), Cxcl11(53.33x), Ifng(4.71x), Stat1(14.48x), Jak1(1.40x)
γ_c	IL15(2.24x), JAK3(2.10x), IL2RG(2.08x), IL2RB(1.98x), IL15RA(1.60x), IL21R(1.84x), IL2RA(1.12x), IL7(2.20x), IL7R(1.61x)	Il15(0.80x), Jak3(1.39x), Il2rg(2.94x), Il2rb(3.14x), Il15ra(1.40x), Il21r(1.69x), Il2ra(1.57x), Il7(1.99x), Il7r(2.39x)
Other	ST8SIA4(2.94x), GPR65(2.60x), GLIPR1(2.29x), IKZF1(2.29x), CD274(2.17x), SAMD9L(2.16x), LCP1(2.15x), SASH3(2.09x), ATP8B4(2.07x)	St8sia4(1.86x), Gpr65(3.25x), Glipr1(2.33x), Ikzf1(2.34x), Cd274(8.41x), Samd9l(4.37x), Lcp1(2.25x), Sash3(2.76x), Atp8b4(3.44x)

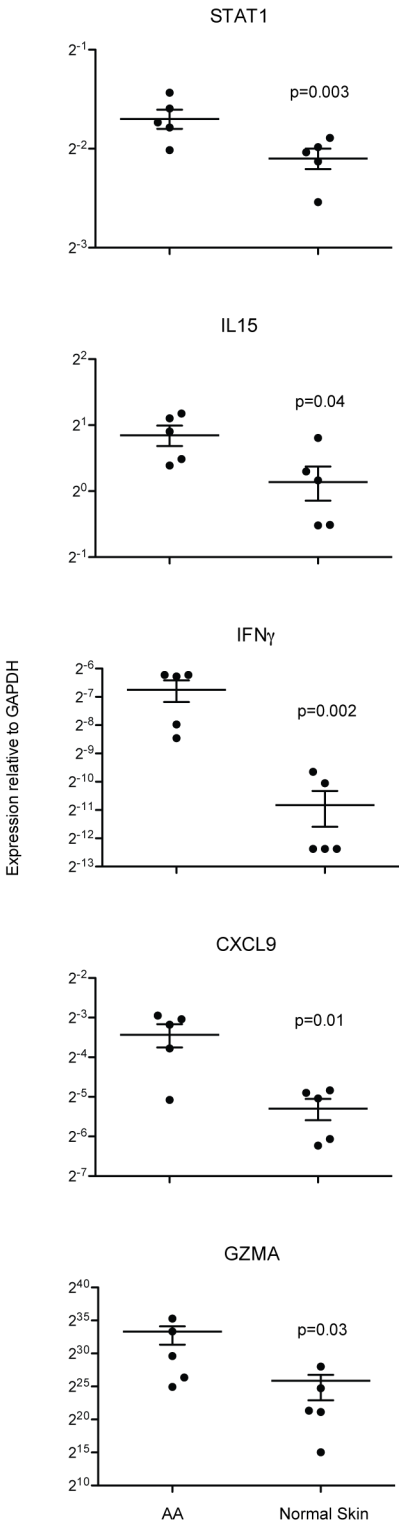


Supplementary Figure 3. Validation of mouse RNA expression studies.

To determine the expression signature of C3H/HeJ mouse skin affected with alopecia areata, lesional skin was isolated from three affected female mice and three unaffected aged-matched controls. Total and small RNAs were isolated from whole skin and biotin-labeled cRNA was generated through *in vitro* transcription, followed by hybridization to the Affymetrix Mouse 430

Nature Medicine: doi:10.1038/nm.3645

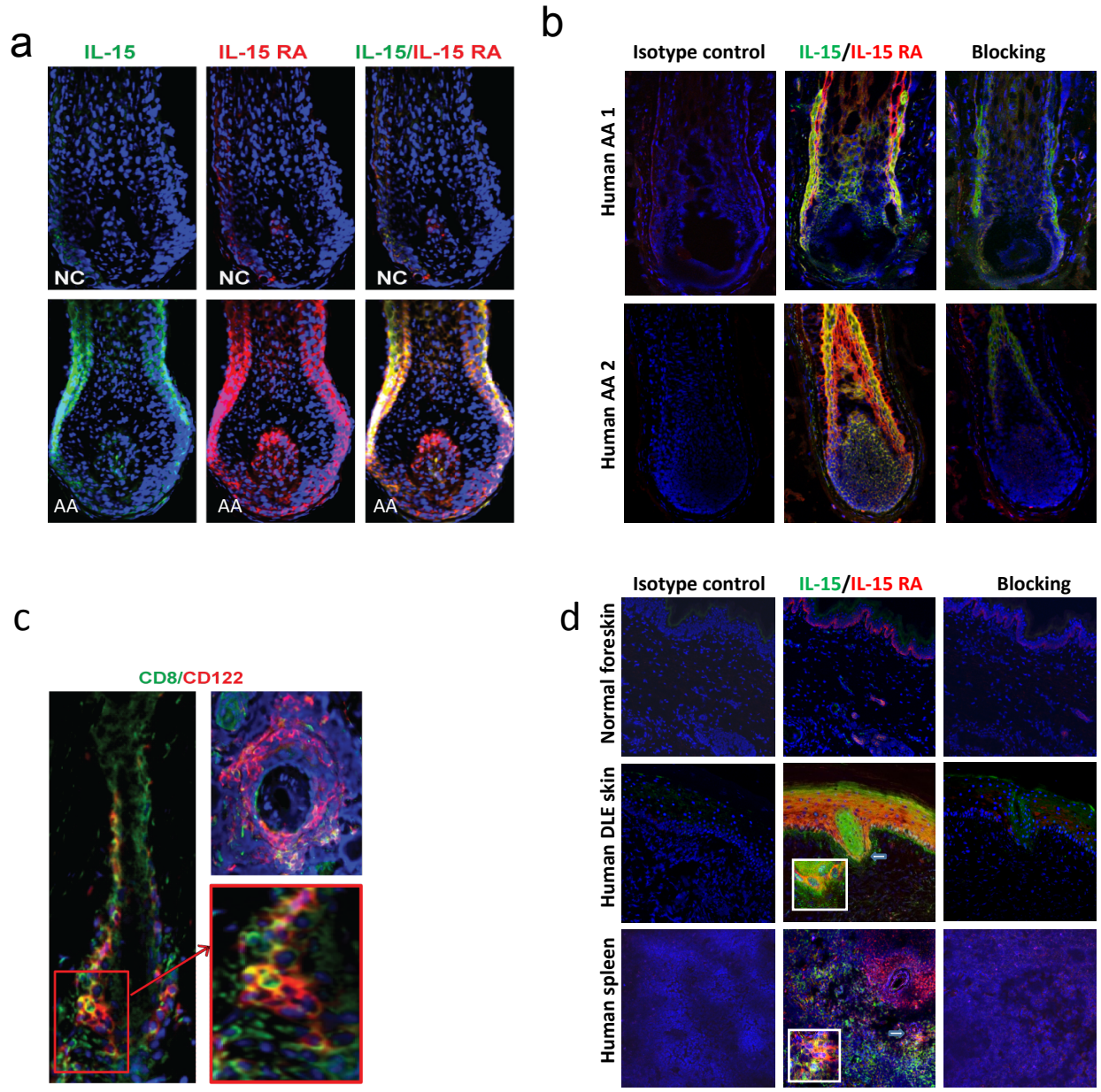
2.0 Genechip. Data analysis was done as outlined in the Methods. Top panel, Representative list of differentially expressed genes among human AA and C3H/HeJ AA reveals shared inflammatory pathways, in particular IFNγ pathway genes, genes representative of CD8 effectors, and a striking γ_c pathway signature. Middle panel, Heatmap depicting the significantly and differentially expressed genes between C3H/HeJ affected and unaffected skin. Bottom panel, Validation by qRT-PCR of several selected immune-related genes from this list whose expression levels are significantly upregulated in AA lesional skin compared to unaffected skin, where each bar represents the average fold change of three independent experiments. UA = unaffected; AA = affected



Supplementary Figure 4. Validation of human mRNA expression studies.

Validation by qRT-PCR of several selected immune-related genes whose expression levels were significantly upregulated by transcriptional profiling in human AA lesional skin (n=5) compared to unaffected skin from healthy controls (n=5).

Nature Medicine: doi:10.1038/nm.3645

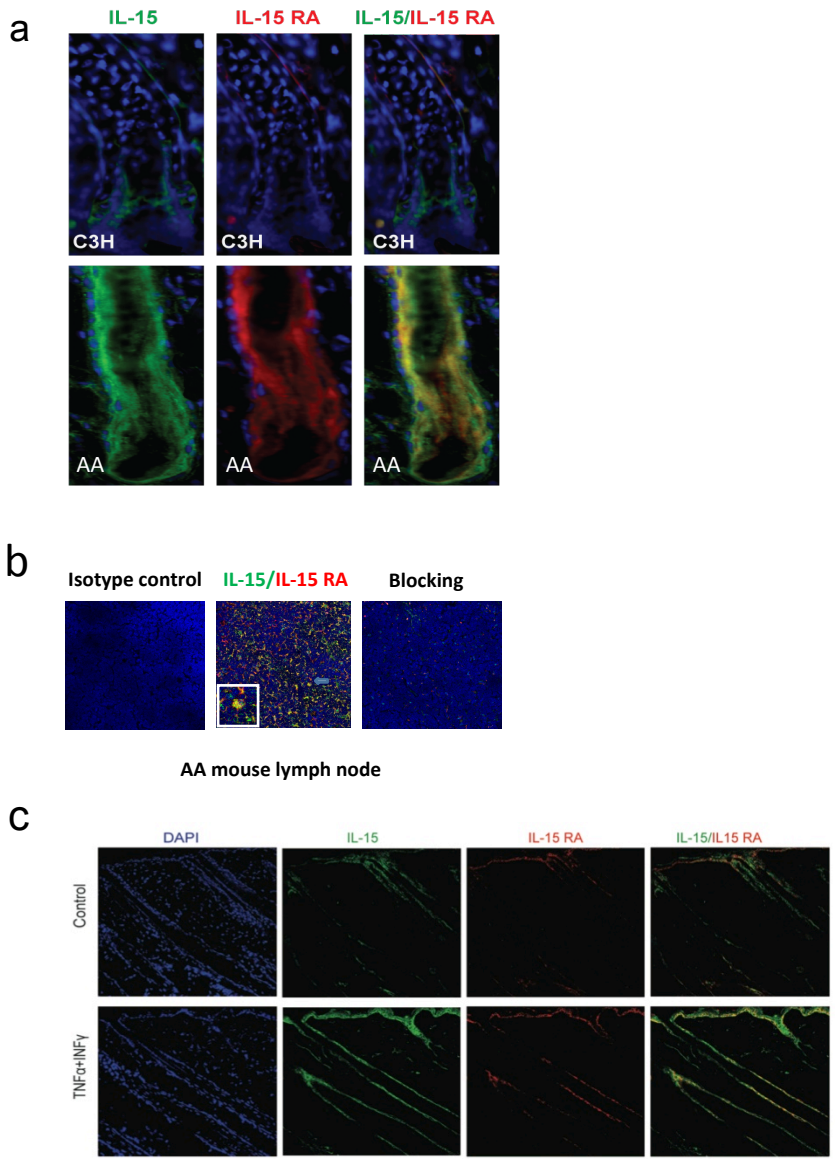


Supplementary Figure 5. IL-15 Staining of Human AA hair follicles.

IL-15 and IL-15RA expression are upregulated in the outer root sheath of human AA (compared with normal controls, NC). a, Human hair follicles stained for IL-15, IL-15Rα, and merged image. b, Frozen 5μM human hair follicle sections from two patients with AA, stained with isotype control, anti-IL-15 and anti-IL-15RA without and with blocking reagents as indicated above. c, Hair follicle infiltrates contain $CD8^+$ T cells that co-expressed CD122 (IL-15Rβ). d, As a positive control, human lesional skin from a patient with discoid lupus was used. Discoid lupus skin sample and spleen tissue show positive

Nature Medicine: doi:10.1038/nm.3645

staining of IL-15 and IL-15Rα, while normal foreskin does not. Isotype control: normal rabbit IgG control for rabbit anti-IL15 antibody, or normal goat IgG for goat anti-IL15RA antibody. Blocking: IL-15 protein plus anti-IL15 antibody, or IL-15RA peptide plus anti-IL15 RA antibody.

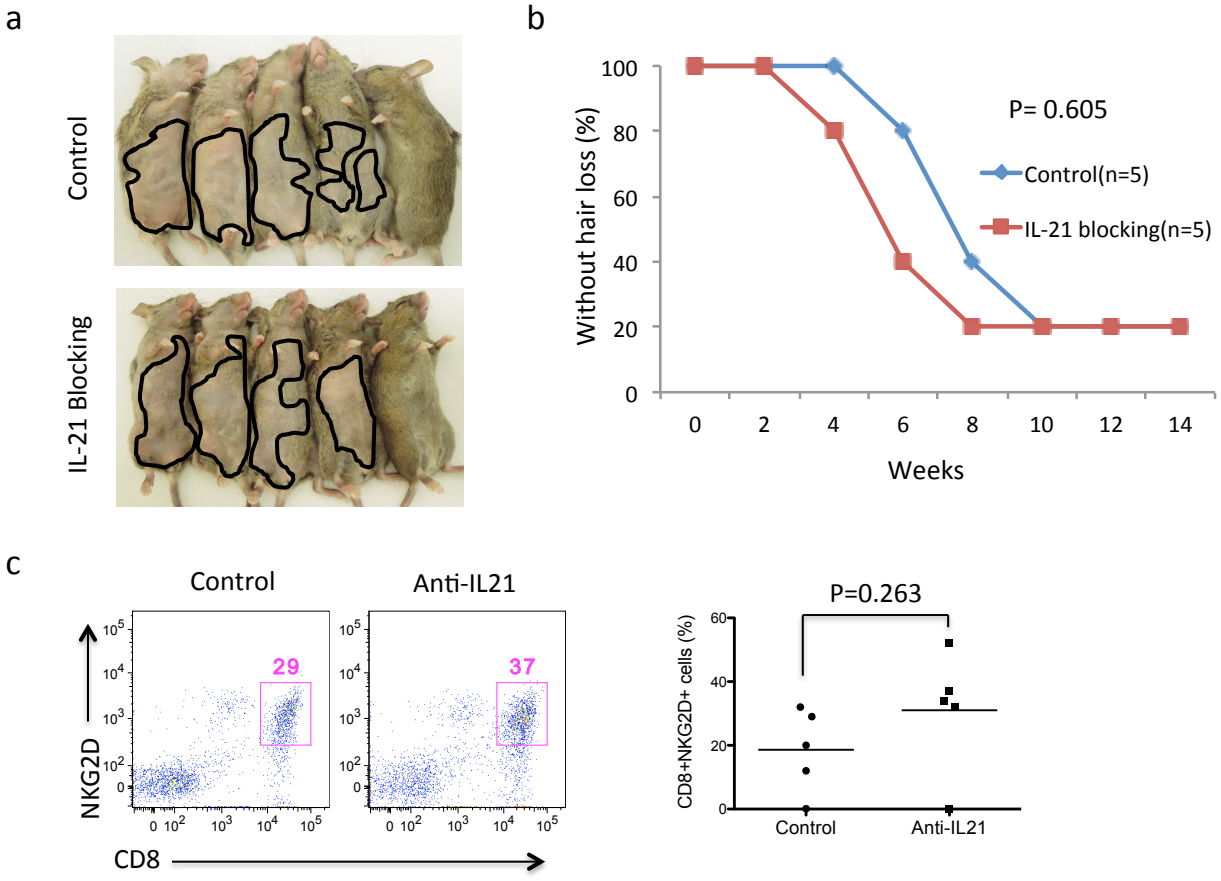


Supplementary Figure 6. IL-15 Staining of mouse AA hair follicles.

a, IL-15 and IL-15Rα expression are upregulated in the outer root sheath of mouse AA (compared with non-alopecic controls, C3H). b, The positive staining of IL-15/IL-15Rα in alopecia mouse cutaneous lymph nodes is inhibited by blocking reagents. Isotype control: normal rabbit IgG control for rabbit anti-IL15 antibody, or normal goat IgG for goat anti-IL15RA antibody. Blocking: IL-15 protein plus anti-IL15 antibody, or IL-15RA peptide plus anti-IL15Rα antibody. c, IL-15 and IL-15Rα expression is induced after stimulation with IFN-γ/TNF-α. Mice received intradermal injections of 100 ul PBS as

Nature Medicine: doi:10.1038/nm.3645

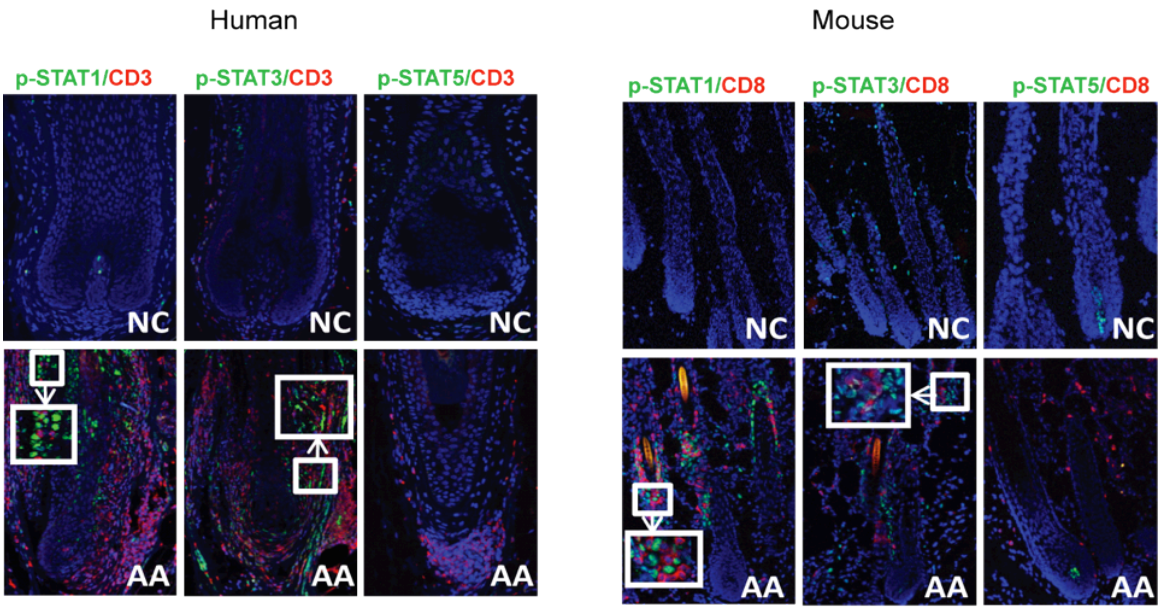
control or 1000ng IFNγ and 250ngTNFα in 100ul PBS, and were sacrificed 72 hours later. Skin sections were stained for IL-15 and IL-15RA.



Supplementary Figure 7. IL-21 blockade fails to prevent the onset of Alopecia Areata in C3H/HeJ mice.

C3H/HeJ grafted mice were treated from the time of grafting with 300 mg anti-IL-21 antibody (Ebioscience, Clone FFA21) injected intraperitoneally three times a week for 14 weeks. (a) The onset of alopecia is not inhibited by anti-IL-21 treatment in grafted C3H/HeJ mice. (b)Time course of the onset of hair loss in control mice and anti-IL-21 treated mice was shown as weeks after grafting. (c) The frequency of $CD8^{+}NKG2D^{+}$ T cells in skin of anti-IL-21 treated mice were not significantly different compared to control mice.

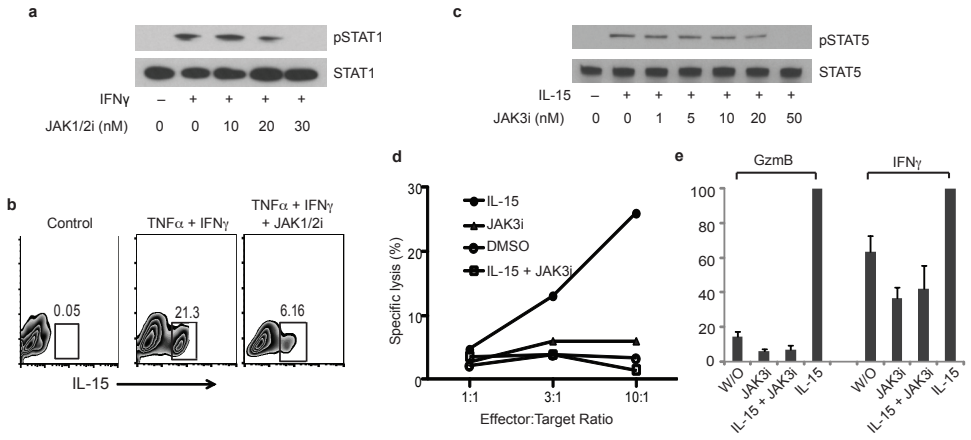
Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 8. STAT expression in human and mouse hair follicles.

Skin sections from human normal controls (top/left panels) or human AA hair follicles (bottom/left), were costained for CD3 and a phosphorylated STAT molecule as indicated above. Mouse skin sections from unaffected C3H/HeJ mice (top right) or alopecic C3H/HeJ mice (bottom right) were costained for CD8 and a phosphorylated STAT molecule as indicated above.

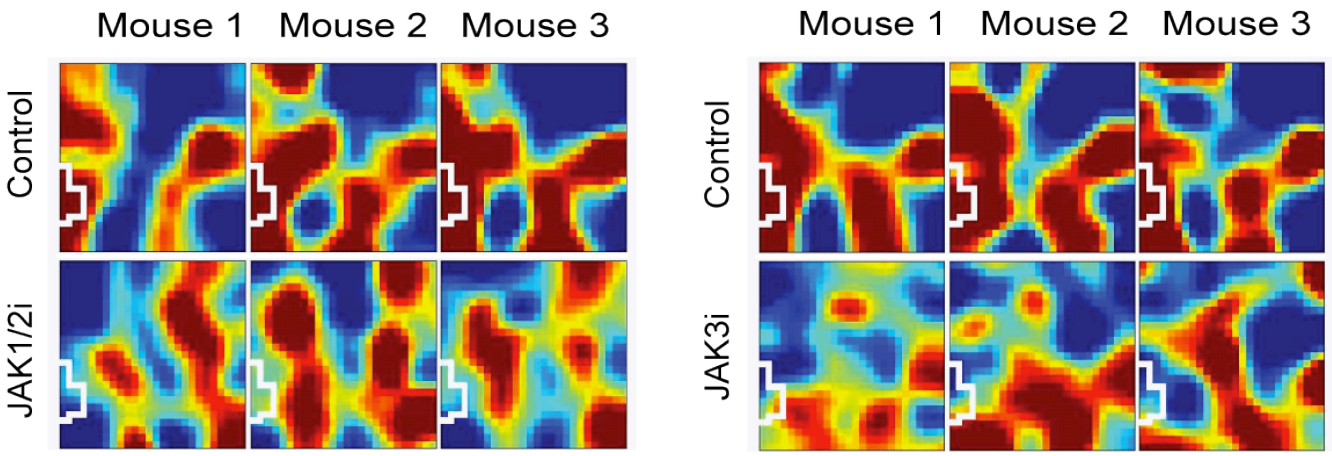
Nature Medicine: doi:10.1038/nm.3645



Supplemental Figure 9. JAK inhibitors diminish responses to IFN-γ and IL-15.

a, JAK1/2i (ruxolitinib) inhibits IFNγ-induced Stat1 activation and b, IL-15 production in dermal sheath cells. c, JAK3i (tofacitinib) inhibits IL-15-induced Stat5 activation in T cells. d, JAK3i inhibits IL-15-induced LAK cell cytotoxic function. e, JAK3i inhibits IL-15-induced LAK cell granzyme B expression and IFNγ production.

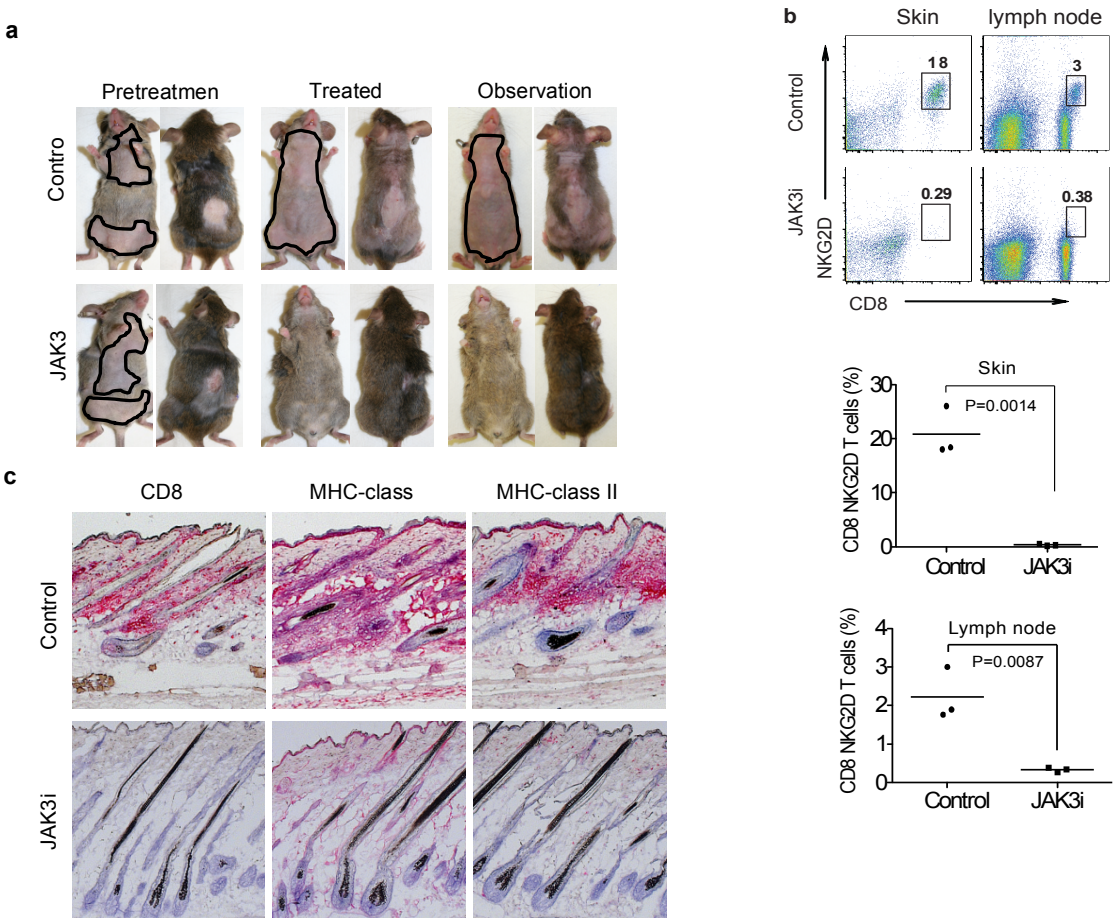
Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 10. GEDI analysis from prevention studies.

GEDI analysis was performed on gene profiles from JAK1/2i- or JAK3i- treated mice. A subjectively identified region of metagenes (white outline) that correlated with either the clinical appearance or resolution of alopecia identified many CTL and IFN related genes.

Nature Medicine: doi:10.1038/nm.3645



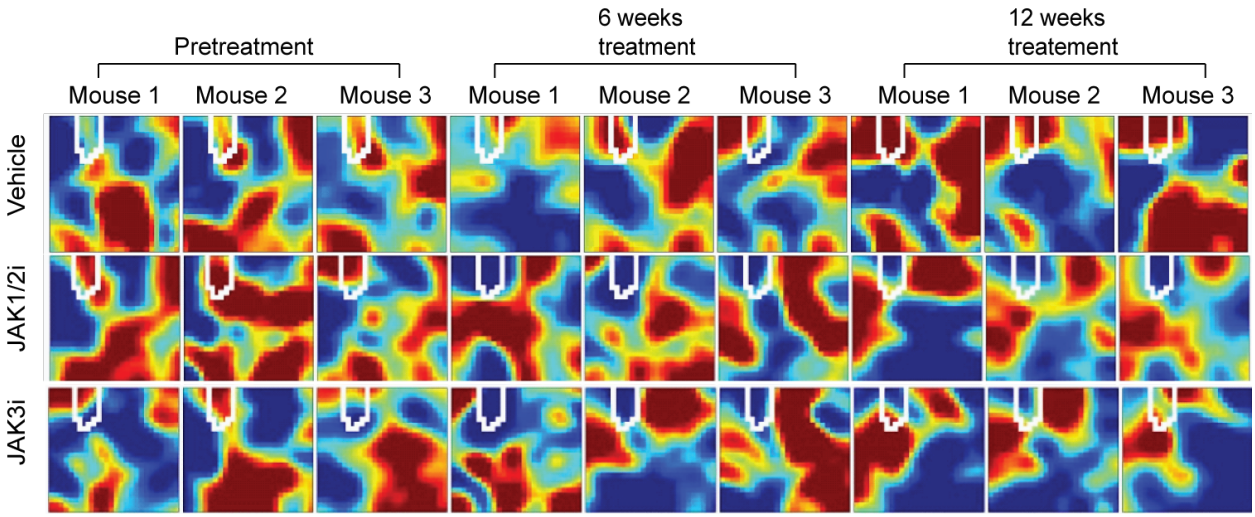
Supplementary Figure 11. Systemic treatment of AA mice with JAK3 inhibitor.

C3H/HeJ mice with long-standing alopecia areata were treated with tofacitinib with Alzet osmotic mini-pumps (pumps, model 2004, Durect Corporation) implanted subcutaneously on the back of each mouse to deliver vehicle (poly(ethylene glycol) (PEG)300) or vehicle containing JAK3i tofacitinib (Abmole) at 15 mg/kg/day for 12 wks. **a,** Alopecia areata reversal was complete on both the back and belly, although the rate of hair regrowth was slower than with topical administration. **b,** Flow cytometric analysis of skin and cutaneous lymph node populations shows elimination of the $CD8^{+}NKG2D^{+}$ T cell population in treated mice (n=3 per group). **c,** Immunostaining of skin from mice

Nature Medicine: doi:10.1038/nm.3645

treated with tofacitinib or placebo demonstrates elimination of CD8 infiltration and MHC I and II upregulation in tofacitinib treated mice.

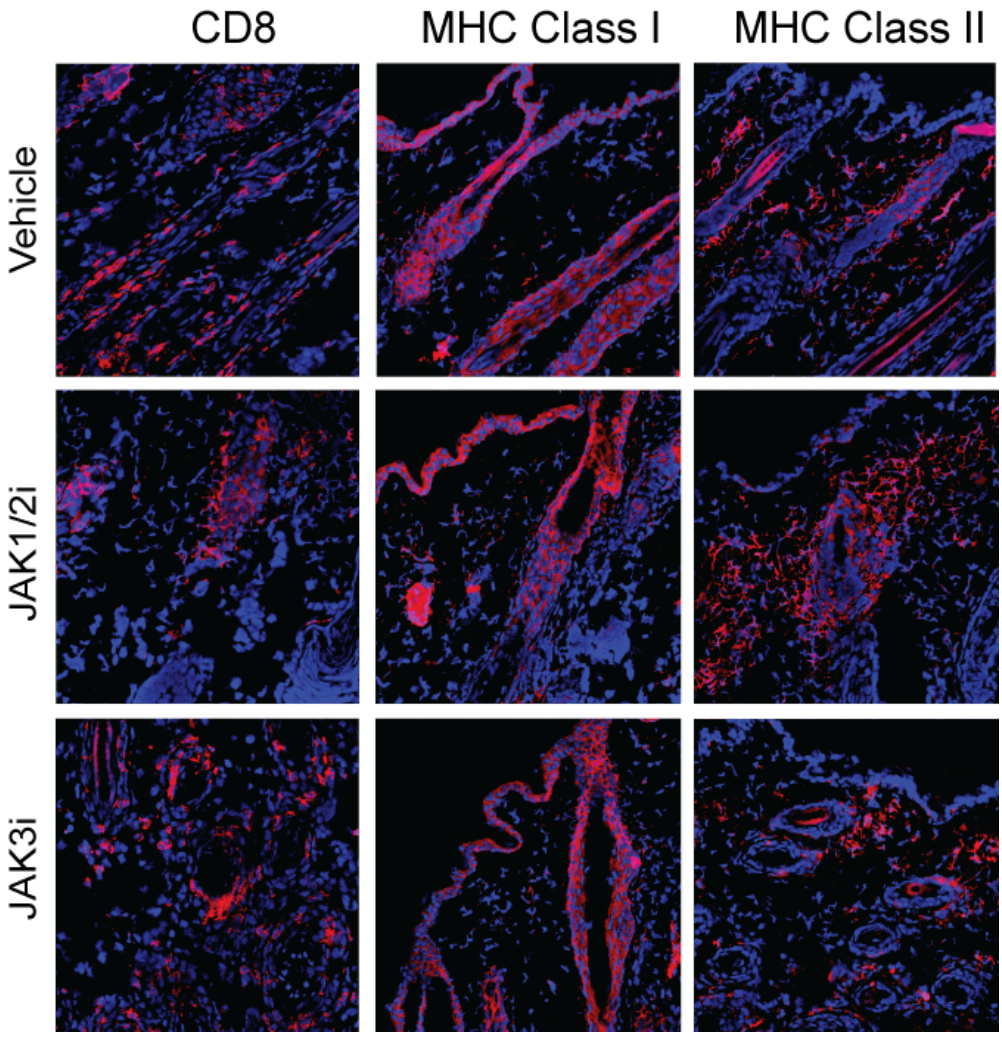
Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 12. GEDI of topical treatments.

GEDI analysis was performed on C3H/HeJ mice before treatment (left 3 rows of panels), after 6 weeks of treatment (middle 3 rows of panels), or after 12 weeks of treatment (right 3 rows of panels. Treatments included topical formulations of ruxolitinib (middle row), tofacitinib/JAK3i (bottom row) compared to vehicle control (top row).

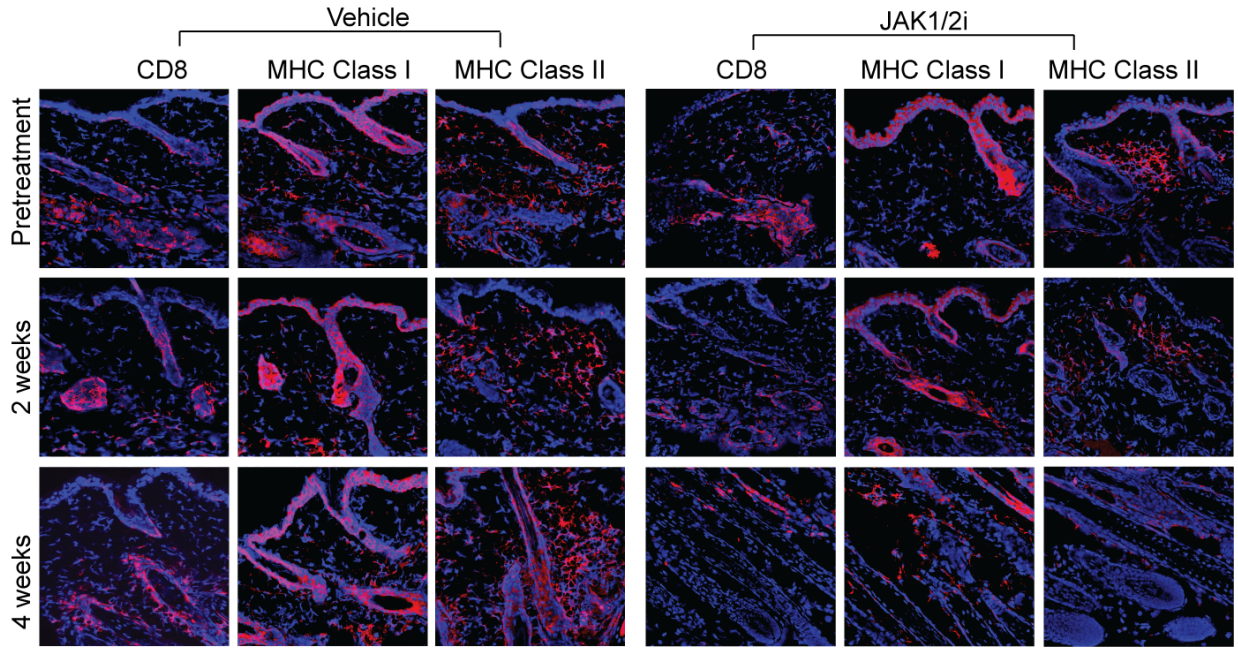
Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 13. Immunofluorescence staining of untreated abdominal skin.

Ventral skin sections from mice treated with topical formulations of vehicle control (top row), ruxolitinib/JAK1/2i (middle row), or tofacitinib/JAK3i (bottom row) on their dorsal surfaces were stained with CD8 (left column), MHC Class I (middle column), or MHC Class II (right column). All sections were costained with DAPI.

Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 14. Immunofluorescence staining of skin at early timepoints during topical treatment.

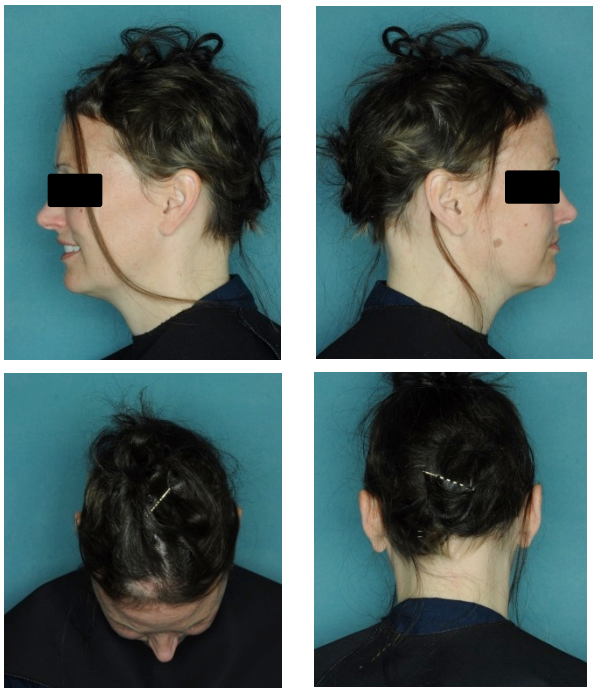
Skin sections from mice treated with topical formulations of vehicle control (left columns) or ruxolitinib/JAK1/2i (right columns) were stained with CD8, MHC Class I, or MHC Class II as indicated. Skin sections were stained prior to starting treatment (top row), or after 2 (middle row) or 4 weeks (bottom row) of treatment. All sections were costained with DAPI.

Nature Medicine: doi:10.1038/nm.3645

Subject #1

A. Baseline **B. 12 Weeks Ruxolitinib Treatment**

C. 20 Weeks Ruxolitinib Treatment

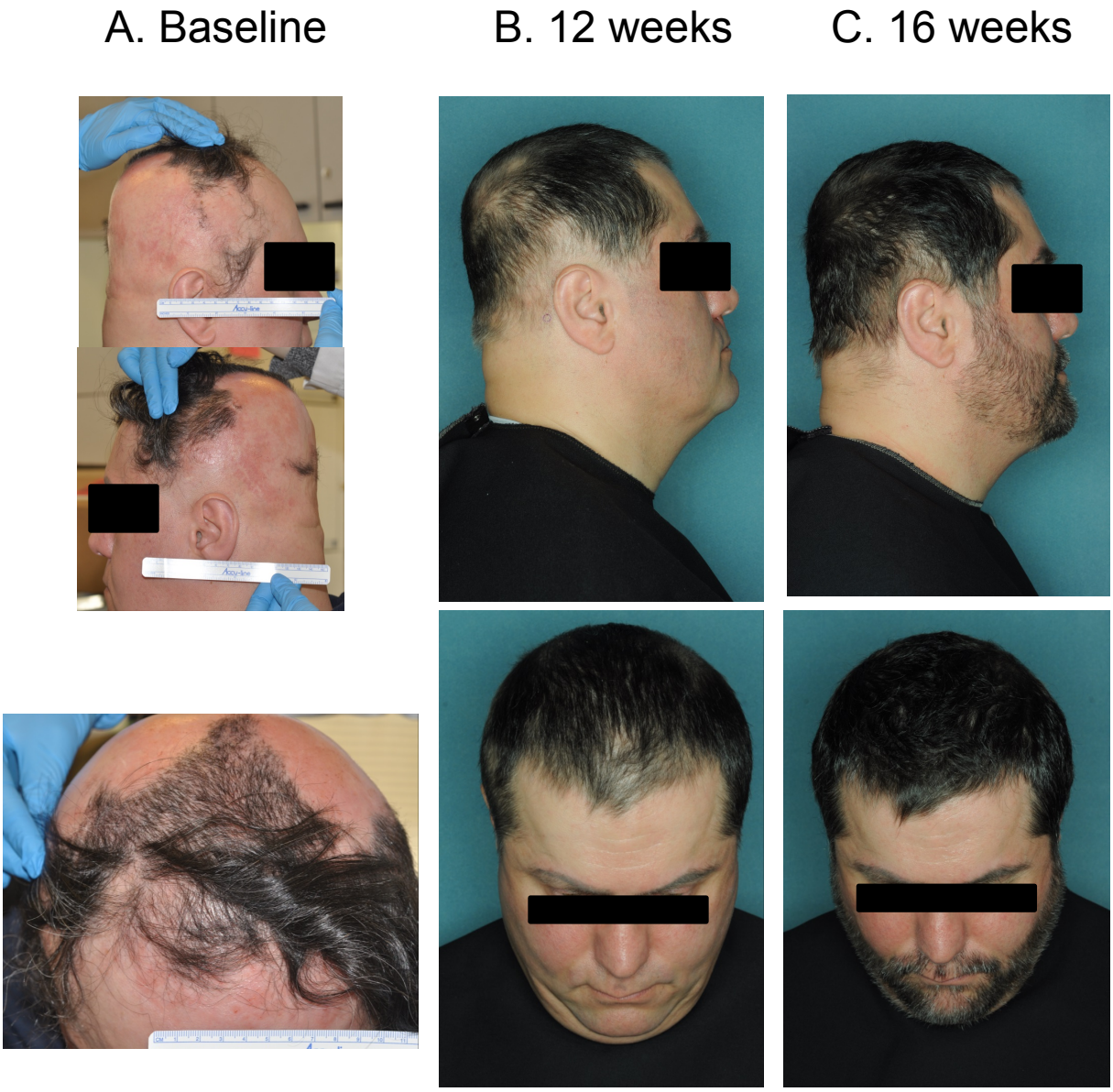


Supplemental Figure 15. Patient #1 response to ruxolitinib.

Patient #1 at baseline (A) and after 12 weeks (B) and 20 weeks (C) of treatment. Near complete hair re-growth is seen by 20 weeks. Patient #2, data not shown, also exhibited near complete hair re-growth by 12 weeks, except in the peri-auricular areas.

Nature Medicine: doi:10.1038/nm.3645

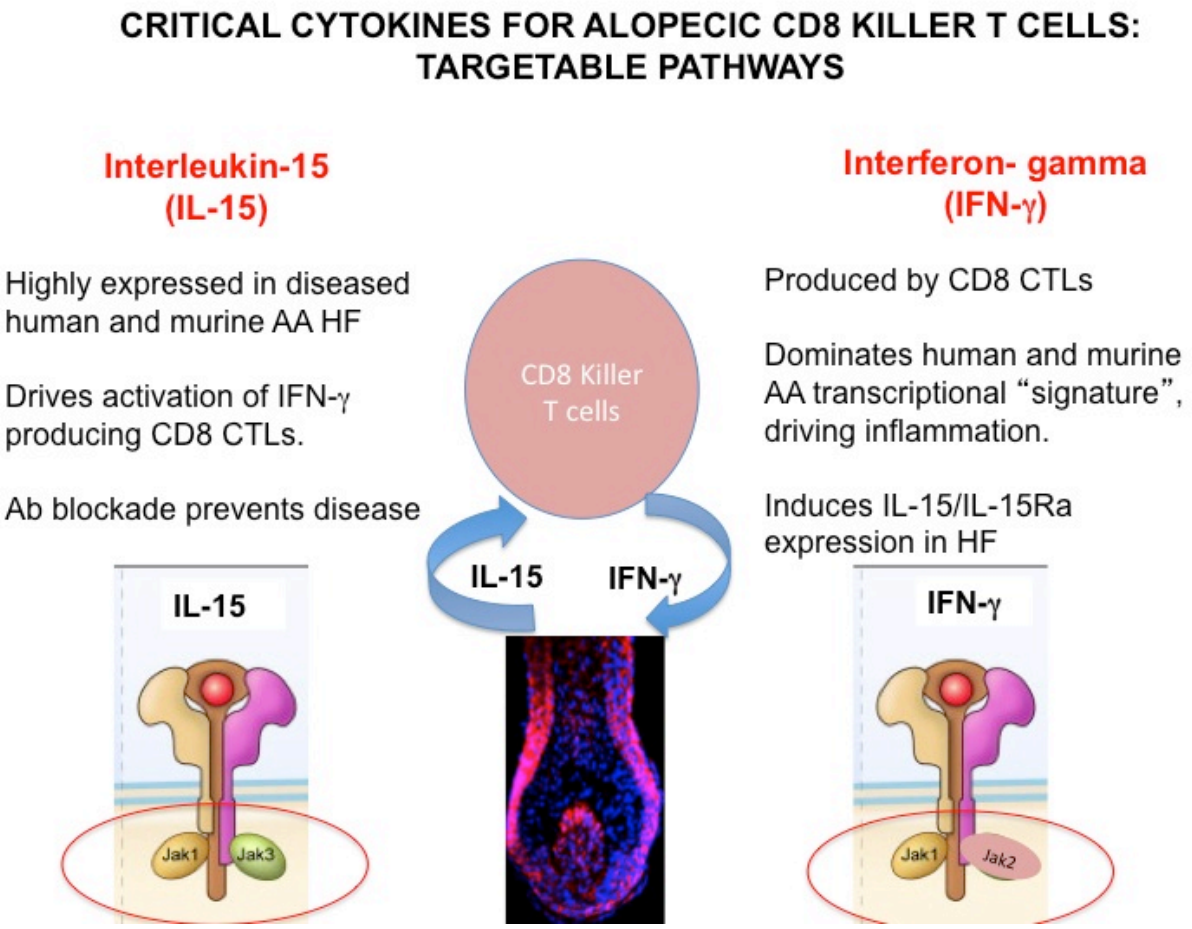
Subject #3



Supplemental Figure 16. Patient #3 response to ruxolitinib.

Patient #3 at baseline (a), after 12 weeks (b) and after 16 weeks (c) of treatment. Patients enrolled in this pilot study are anticipated to be treated until maximal response or up to 6 months of treatment, followed by observation off drug for an additional 3 months to assess response durability.

Nature Medicine: doi:10.1038/nm.3645



Supplementary Figure 17. Schematic representation of our model of the IL-15/IFNγ axis in alopecia areata and its targeting using JAKis.

Interferon gamma and IL-15/IL-15Ra cooperate to induce a Type I cytotoxic response. CD8 T cells produce IFNγ which breaks the immunological privilege of the hair follicle, upregulating antigen presentation (MHC I and II), inducing inflammatory chemokines (CXCL9-11), induction of NKG2DLs and upregulation of IL-15/IL-15Ra by the hair follicle. In turn, this leads to the further recruitment and activation of effector $CD8^{+}NKG2D^{+}$ T cells.

Nature Medicine: doi:10.1038/nm.3645

Chapter IV.4

Functional genomics at the *ULBP3/6* locus identifies a role for CTCF variants in alopecia areata

Gina M. DeStefano[1], Lynn Petukhova[2,3], Esther Drill[2,4], Zhenpeng Dai[2], Li Bian[2], Raphael Clynes[2,5], and Angela M. Christiano[1,2]*

Departments of Genetics and Development[1], Dermatology[2], Epidemiology[3], Biostatistics[4], and Medicine[5], Columbia University, New York, NY, 10032

(Manuscript #5, in preparation)

Abstract

Alopecia areata (AA) is among the most highly prevalent human autoimmune diseases, characterized by disfiguring hair loss due to the collapse of immune privilege in the hair follicle. To determine the genetic architecture of AA, the Christiano lab recently performed a Genome-Wide Association Study (GWAS) and identified the *ULBP3/6* locus on chr.6q25.1 (p = 5.9 x 10^{-24}) as the most significantly associated locus outside the *HLA* region. The *ULBP* genes reside within a 180 Kb MHC class I-related cluster and encode ligands for the NKG2D cytotoxic T cell receptor. Importantly, these ligands possess striking upregulation in both human and mouse AA hair follicles, and we showed that CD8+NKG2D+ cells are both necessary and sufficient to induce disease in the AA mouse model, establishing a definitive role for the *ULBP*-NKG2D signaling axis in disease pathogenesis. To identify genetic variants in the *ULBP3/6* locus that are associated with AA, we performed targeted deep resequencing on 124 AA patients and uncovered a number of rare, non-coding variants enriched on GWAS risk haplotypes. Interestingly, we identified three novel variants that reside within *ULBP6* regulatory elements that coincide with CTCF ChIP-seq binding sites. These variants fall within paired-end tags from an ENCODE CTCF-mediated Chromatin Interaction with Paired-End Tags (ChIA-PET) experiments, suggesting this region engages in long-range interactions. *In vitro, ULBP6* AA rare variants decreased reporter activity and abrogated CTCF-mediated *ULBP6* repression, and *in vivo*, we observed CTCF binding enrichment at *ULBP6* by ChIP-PCR on human scalp dermal fibroblasts. Our findings identify a role for CTCF variants at the *ULBP3/6* locus in association with AA, suggesting that long-range regulatory interactions may control gene expression in the hair follicle.

Introduction

Alopecia areata (AA) is an autoimmune disease characterized by disfiguring hair loss that affects approximately 5.3 million people. An underlying genetic basis for AA is supported by the observed heritability in first-degree relatives (10-fold increased risk), twin studies (42-55% concordance in monozygotic twins) and family-based linkage studies (VAN DER STEEN *et al.* 1992; MCDONAGH AND TAZI-AHNINI 2002) (JACKOW *et al.* 1998; MARTINEZ-MIR *et al.* 2003), though the key pathogenic risk variants underlying this complex disease remained elusive.

The Christiano lab conducted an initial Genome-Wide Association Study (GWAS), identifying eight regions that clustered into discrete linkage disequilibrium (LD) blocks and surpassed genome-wide statistical significance, including the *HLA (MICA/MICB), ULBP3/6, CTLA4, IL2RA, IL2/IL21, IKZF4 (Eos), STX17* and *PRDX5* loci (PETUKHOVA *et al.* 2010). To expand gene discovery, we recently conducted a meta-analysis GWAS and replication expanding our cohort up to 3,253 cases and 7,543 controls, which identified two novel loci and increased the genome-wide significance of our previous findings (BETZ *et al.* 2015). Remarkably, our GWAS/meta-GWAS results pointed to the *ULBP3/6* region on chr.6q25.1 ($p = 5.9 \times 10^{-24}$) as the most significantly associated locus outside the *HLA* region.

The *ULBP* (cytomegalovirus UL16-binding protein) genes reside in a 180-kilobase MHC class I-related cluster on human chromosome 6q25.1, and encode multiple ligands for the activating cytotoxic T cell receptor, NKG2D (CAO AND HE 2004). Our GWAS and subsequent replication are the first to implicate the *ULBP* gene cluster in any human disease, although allelic variants and polymorphisms in other NKG2DLs

(MICA/MICB) have been previously associated with AA (BARAHMANI *et al.* 2006; ITO *et al.* 2008).

The NKG2D receptor is expressed on CD8+ cytotoxic T cells and NK cells, and becomes activated in the presence of a stressed cell upregulating *ULBP* ligand expression as a "danger signal" (CAILLAT-ZUCMAN 2006). NKG2D-medated cytotoxicity has been well described in the context of viral infection, autoimmunity as well as cancer (GONZALEZ *et al.* 2006). The Christiano lab previously demonstrated a marked upregulation of the NKG2D ligands, ULBP3 and MICA on human AA lesional hair follicles, and showed that NKG2D+CD8+ T cells infiltrate the base of the hair follicle during active disease. In the alopecia areata mouse model (C3H/HeJ strain) that recapitulates many pathological features of adult AA in humans, we observed an upregulation of the NKG2D ligands, H60 and Rae-1 (analogous to human MICA and ULBP), and demonstrated that cytotoxic CD8+NKG2D+ T cells are both necessary and sufficient for the induction of hair loss and disease pathogenesis (PETUKHOVA *et al.* 2010; XING *et al.* 2014) (BARAHMANI *et al.* 2006; ITO *et al.* 2008). These immunological findings and the strong genetic association of *ULBP3/6* with human AA provide a definitive role for the *ULBP*-NKG2D signaling axis, yet the precise pathogenic NKG2DL variants contributing to autoimmunity in the hair follicle remain unknown.

There are several well-characterized immune-related loci that are transcriptionally regulated in a spatiotemporal manner by long-range chromosomal interactions (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS *et al.* 2013; SHARAF *et al.* 2014), including IFNG, class II cytokine receptor (CIICR) genes, Th2 genes including *IL4, IL5,* and *IL13*, and the *HLA* region

encoding the Major Histocompatability Complex (MHC) molecules that are transcriptionally regulated in a spatiotemporal manner by long-range chromosomal interactions (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS *et al.* 2013; SHARAF *et al.* 2014). Many of these genes reside within clusters, which imply genetic structures governed by long-range interactions. The chromatin remodeling insulator protein CTCF has been described in several contexts to facilitate these interactions during lineage commitment, the generation of antigen diversity, as well during an immune response (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS *et al.* 2013; SHARAF *et al.* 2014). Considering the strong association of *ULBP3/6* with AA and predominantly intergenic location of GWAS SNPs, we postulated that non-coding Single Nucleotide Variants (SNVs) contribute to dysregulated transcription factor binding, *ULBP* regulatory activity, and/or long-range interactions disrupting *ULBP3/6* repression in the AA hair follicle.

Here, we identified and characterized novel and rare susceptibility variants by performing targeted deep resequencing the *ULBP3/6* locus. Using a robust functional genomics approach combined with *in vivo* and *in vitro* analyses, we identified the *ULBP6* candidate region as a site for CTCF binding enrichment in human scalp dermal fibroblasts, and demonstrated that AA rare susceptibility variants not only disrupt regulatory activity, but also abrogate CTCF-mediated repressive effects on the *ULBP6* sequence. Moreover, the novel, rare variants are predicted to create new TFBS for activating proteins in immune signaling pathways. Our findings point to a novel role for CTCF and long-range interactions in the pathology of AA, providing insight into the mechanisms contributing to dysregulated *ULBP3/6* expression in the AA hair follicle.

Results

Targeted deep re-sequencing of the *ULBP3/6* locus and functional genomics studies reveal novel rare, non-coding variants in CTCF binding sites

The initial GWAS revealed eight genomic regions that are significantly associated with AA and cluster into discrete linkage disequilibrium blocks (PETUKHOVA *et al.* 2010). Since GWAS tag SNPs are intended to capture the majority of common genetic variation at a particular region, they usually represent a very small portion of the variants and provide very little insight about rare variants. Therefore, it is necessary to perform targeted deep resequencing to identify both rare and common variants that are driving the genetic association of the tagSNPs. We prioritized these regions for targeted enrichment based on our GWAS LD peaks and our functional studies in the C3H/HeJ mouse model demonstrating a crucial role for the NKG2D signaling pathway in AA pathology (XING *et al.* 2014).

To define the candidate region at the *ULBP3/6* locus to be resequenced, we initially performed conditional analysis and determined that the 13 tag SNPs identified from the initial GWAS formed two statistically independent risk haplotypes (color coded blue and yellow in Figure 1A), encompassed within a 72 Kb region including both *ULBP3* and *ULBP6*. We next designed a RainDance primer library to identify variants carried on these GWAS defined risk haplotypes and deep sequenced the entire 72 Kb region in 122 AA GWAS cases (Figure 1A). We identified 7 exonic variants, 48 intronic variants, and 364 intergenic variants. To identify rare variants potentially enriched in our cohort of AA patients, we selected all variants new or reported to be rare in public databases (f<0.01; n=129) that are found in 3 or more patients (n=29), which yielded a

total of 189 potentially causal variants (29 rare variants enriched on GWAS haplotypes and 160 common variants associated with risk haplotypes). We next performed functional annotations of the non-coding variants using Encyclopedia of DNA elements (ENCODE) and RegulomeDB to prioritize variants that may disrupt transcription factor binding sites (TFBS) and regulatory function in human keratinocytes and dermal fibroblasts, which are the target cells under autoimmune attack in AA.

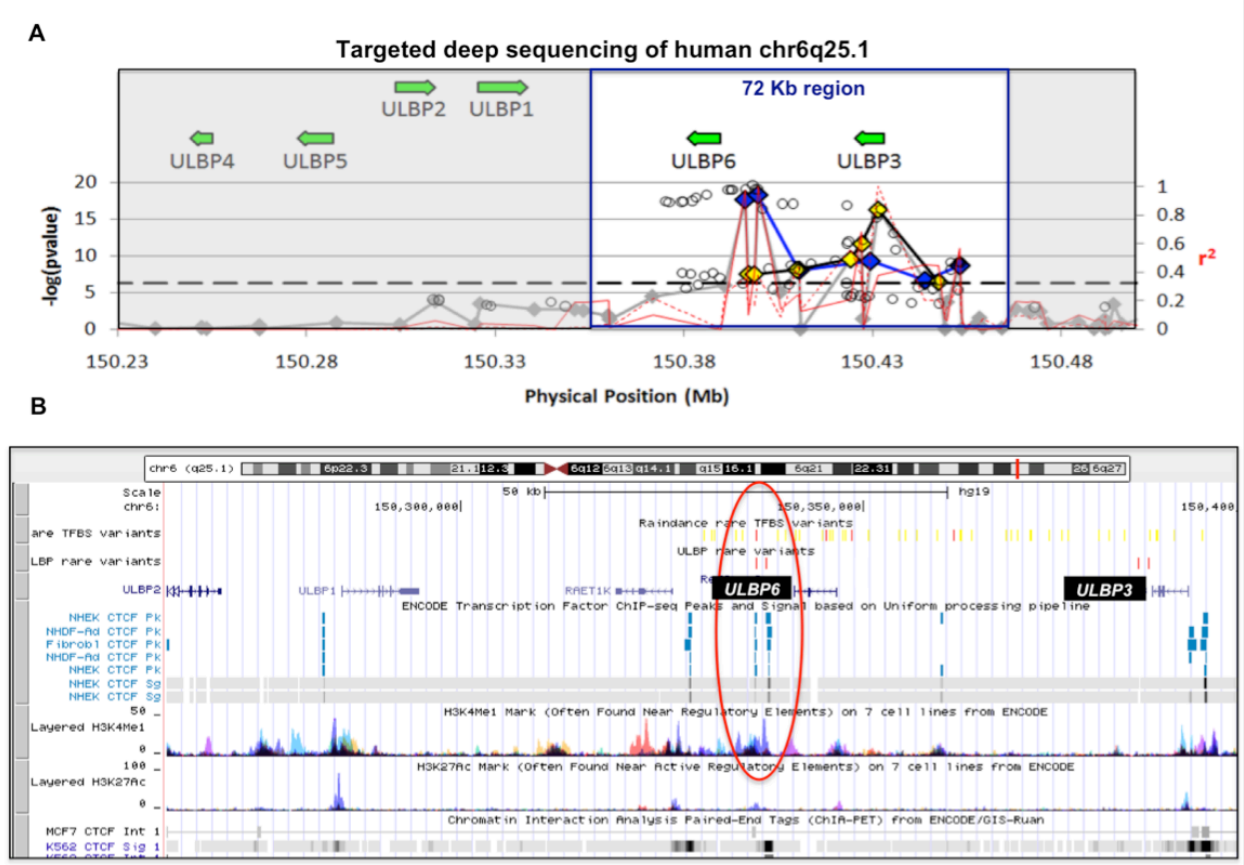


Figure 1. Targeted deep re-sequencing of the *ULBP3/6* locus and functional genomic studies in AA patients reveal rare and novel, non-coding variants in CTCF binding sites. A. Targeted region (72 Kb) for deep resequencing performed on 124 AA patients using the RainDance sequencing technology. Blue and yellow lines indicate two distinct haplotypes with highly significant GWAS tag SNP association. Left Y axis represents the –log(p value) and right Y axis represents the coefficient approximations to LD (r^2). Typed SNPs are plotted as diamonds, and imputed SNPs are plotted as circles. **B**. UCSC Genome Brower (hg19) view of chromosome 6q25.1 and custom tracks displaying rare variants identified from targeted deep sequencing (red) superimposed onto ENCODE CTCF ChIP-seq data from human keratinocytes and dermal fibroblasts, and layered H3K27Ac (a mark associated with open chromatin) and H3K4me1 (a mark associated with regulatory elements).

We identified 21 rare variants that fall within ENCODE defined regulatory regions and areas of chromatin accessibility (overlapping with DNAse I hypersensitive sites and H3K27Ac peaks), and 37 rare variants in known TFBS based on ChIP-seq data in human keratinocytes and fibroblasts (Figure 1B). We discovered a region with potential regulatory function at the *ULBP6* locus harboring a cluster of rare and novel variants (not yet reported in public databases) (Figure 1B). Interestingly, three variants (G→C [chr6:150,336,528], G→T [chr6:150,336,541], and T→C [chr6:150,336,614]) reside within intergenic CTCF binding sites ~5 Kb downstream from the 3' end of *ULBP6* with ChIP-seq evidence for CTCF binding (Figure 1D). At least one of these novel variants was present in 9 independent samples in our cohort of 119 samples, generating a collapsed allele frequency of 3.8%. In addition, at a site ~1.1 Kb downstream of the CTCF variants, we found an enrichment of a rare variant (rs113050885) that also falls within a SUZ12 binding site and is located 28 bp downstream of a highly significant AA GWAS SNP rs9479403 (p-value, 4.4×10^{-18}) and 54 bp upstream of GWAS SNP rs4242284 (p-value 2.1×10^{-08}). Because the two GWAS SNPs are less than 100 bp away from each other, we were able to directly interrogate the Raindance sequence reads and confirmed that in every case, the rare variant rs113050885 lies on a haplotype in phase with the two common risk variants, suggesting that the rare variant is contributing to the association evidence.

CTCF is endogenously bound and enriched at the *ULBP6* candidate region in human scalp dermal fibroblasts

CTCF is a well-characterized repressive insulator best known for its ability to block promoter-enhancer interactions by forming repressive chromatin loops (PHILLIPS AND

CORCES 2009), and previous studies have demonstrated that the genes encoding NKG2D ligands are actively repressed in healthy tissues (CAILLAT-ZUCMAN 2006). Our functional genomic studies at the *ULBP3/6* locus revealed areas of chromatin accessibility (marked by DNAse I hypersensitivity sites), a feature of insulator sites, and CTCF binding by ChIP-seq (ENCODE).

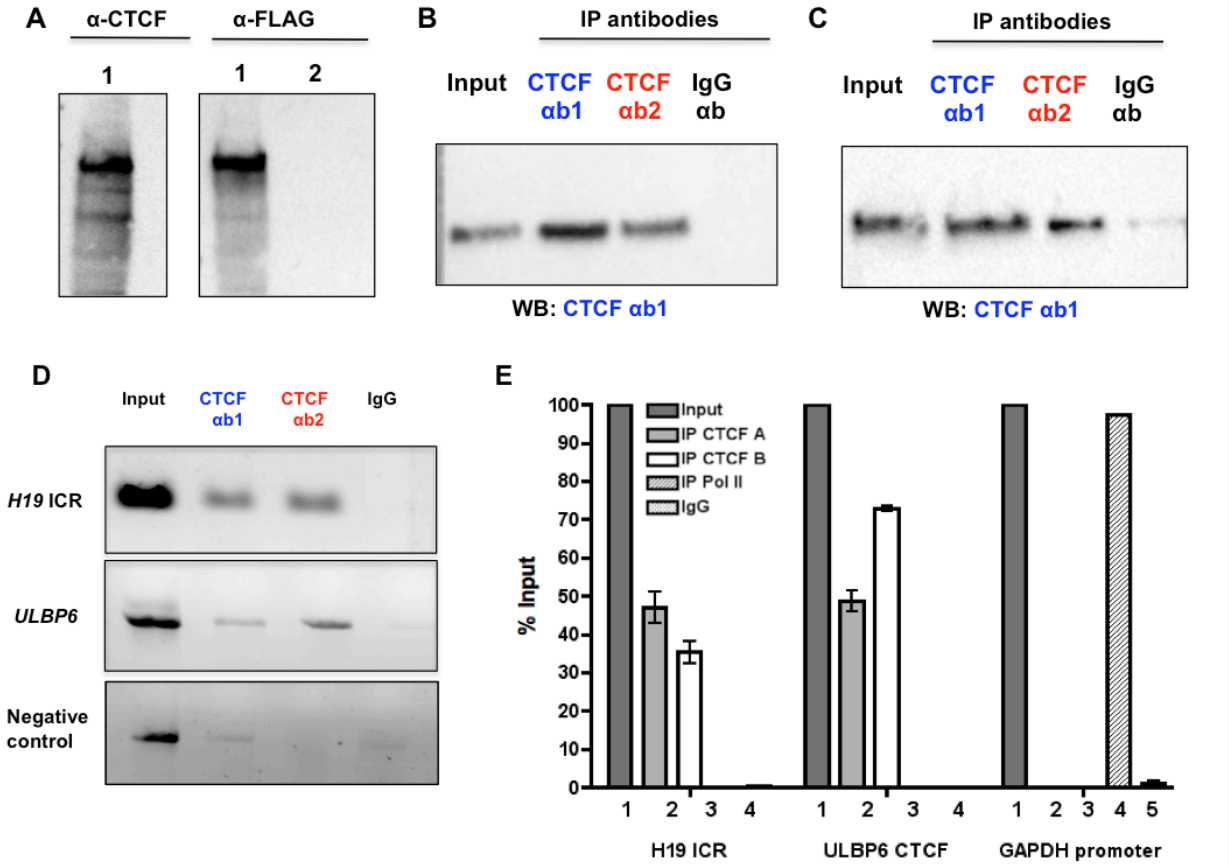


Figure 2. CTCF binding is enriched at *ULBP6* in human scalp dermal fibroblasts. A. Immunoblotting for CTCF and FLAG performed on human dermal fibroblasts transfected with a FLAG-tagged CTCF ORF (1) or untransfected (2) to validate the CTCF antibody. **B, C.** Chromatin Immunoprecipitation (ChIP) performed on primary human scalp dermal fibroblasts (B) and the hTERT DP cell line (C) using two monoclonal CTCF antibodies (Materials and Methods) and IgG as a negative control. Immunoblotting was performed using CTCF antibody 1 (Materials and Methods). **D.** Gel electrophoresis of the ChIP-qPCR results shown in **E**. PCR was also performed on an intergenic region (negative control, chr1q31.3) with no predicted CTCF occupancy. **E.** ChIP-qPCR for the *H19* ICR (positive control for CTCF occupancy), GAPDH (positive control for Pol II occupancy), and the *ULBP6* region. Data are plotted as percent input; bars indicate SD.

To determine whether CTCF is endogenously bound in healthy human hair follicles, we performed Chromatin Immunoprecipitation (ChIP) on two independent dermal fibroblast cell lines derived from the human scalp (Figure 2). Sanger sequencing the *ULBP3/6* candidate region on both cell lines demonstrated that neither carries the AA risk haplotypes in *ULBP3/6* (data not shown). We then determined enrichment for CTCF by immunoblotting using two validated antibodies as well as a FLAG-tagged form of CTCF to demonstrate CTCF antibody recognition (Figure 2A). CTCF immunoprecipitation on dermal fibroblasts overexpressing or lacking the FLAG-CTCF construct, followed by immunoblotting for CTCF and FLAG revealed a ~130 KDa band identical between transfected samples that were probed for α-CTCF and α-FLAG, with no α-FLAG reactivity in the untransfected samples (Figure 2A). Importantly, immunoprecipitation for CTCF using two distinct antibodies produced an intense and identical band on immunoblots that was absent in α-IgG pull downs, demonstrating strong CTCF enrichment (Figure 2B, C).

To validate the efficiency of the ChIP assay on human scalp dermal fibroblasts, we then performed qPCR on the reverse-crosslinked fragmented gDNA for *GAPDH* and the *H19* imprinting control region (ICR), which serve as positive controls for α-Pol II and CTCF occupancy, respectively. We observed enrichment for Pol II at the *GAPDH* promoter (~45% of input) as expected, and CTCF at the *H19* ICR (35-45%), consistent with previous reports (Figure 2D). We next tested for CTCF enrichment specifically at the *ULBP6* candidate region using qPCR, and observed a strong signal (50-70% of input) compared to that of the *H19* ICR, which is only bound by CTCF on the maternal allele (PHILLIPS AND CORCES 2009) (Figure 2D, E). Furthermore, we obtained the same

results using a second set of primers specific to the *ULBP6* candidate region (data not shown), and amplification of a site completely devoid of CTCF occupancy as a negative control (chr1q31.3) yielded no signal (Figure 2E). Taken together, these results clearly demonstrate that CTCF binding is enriched at this locus in healthy human dermal fibroblasts.

Regulatory activity of the *ULBP6* candidate region is reduced in the presence of AA rare, non-coding variants

To gain insight as to the potential regulatory function(s) of the *ULBP6* candidate region harboring the three rare variants, we first cloned a 250 bp fragment spanning the candidate region from control and AA genomic DNA samples into the pGF-mCMV vector (System Biosciences) upstream of a minimal CMV promoter (Figure 3A). We observed a significant increase in reporter activity by ~3 fold ($p < 0.001$) compared to the empty vector, scrambled sequence and a sequence containing no CTCF site (Figure 3B). We next tested the effects of the three rare, non-coding variants on reporter activity *in vitro* using the same approach, where a sequence containing all three variants in the CTCF sites was compared to the wild-type sequence and that containing no CTCF sites (Figure 3B). In the presence of the three AA variants, reporter activity was reduced by ~30% ($p < 0.01$), demonstrating differences in regulatory and *trans* factor binding properties (Figure 3B). To assess whether the decreased signal was attributable to one or more variant and considering that most AA patients with either risk haplotype harbor one rare variant, we cloned the *ULBP6* candidate region harboring the individual rare variants and observed a significant decrease in activity in constructs containing rare variants 2 (G→T) and 3 (T→C), whereas the construct containing rare variant 1

(G→C) displayed activity similar to that of the control sequence (Figure 3C). The levels of rare variants 2 and 3-induced reporter activity were equivalent to the sequence containing all rare variants (Figure 3C), suggesting that these SNVs are producing the greatest effect.

AA rare variants alter CTCF-mediated repressive activity at the *ULBP6* intergenic region

Our previous observations (Figure 3B and C) demonstrate that the AA rare variants reduce the reporter activity compared to that of the control sequence. Given the observation that these variants reside within CTCF binding sites, we next tested the effects of exogenous CTCF on these sequences. Using the pGF-mCMV reporter to test the effects of *trans* binding factors on the *ULBP6* sequence, we first validated the efficiency of the system by cloning in the sequence for the cAMP response element binding protein (CREB), which drives strong reporter activity alone, and is further activated by PMA (phorbol 12-myristate 13-acetate), an activator of Protein Kinase C (PKC) (Figure S2) (MAO *et al.* 2007). We observed a marked increase in luciferase and GFP activity with CREB alone (2500-fold increase) and a nearly three-fold increase upon the addition of PMA to transfected cells ($p<0.05$) (Figure S1). To determine the effects of exogenous CTCF on the control *ULBP6* region, we transfected human dermal fibroblasts with a FLAG-tagged form of CTCF in conjunction with the *ULBP6*-luciferase reporter, and observed a strong repressive effect (65-70% reduction; $p<0.001$) compared to a sequence containing no CTCF binding sites as well as the empty vector (Figure 3D). This finding is consistent with the known role of CTCF as an insulator and repressor, as well as with previous studies demonstrating its repressive effects in long-

range reporter assays (NATIVIO *et al.* 2009). We next tested the consequence of the AA rare variants on CTCF-mediated repressive activity, and observed near abolishment of the repressive effect of CTCF on *ULBP6* reporter activity (Figure 3D). These results suggest that AA rare variants alter the binding affinity of CTCF at the *ULBP6* intergenic regulatory region.

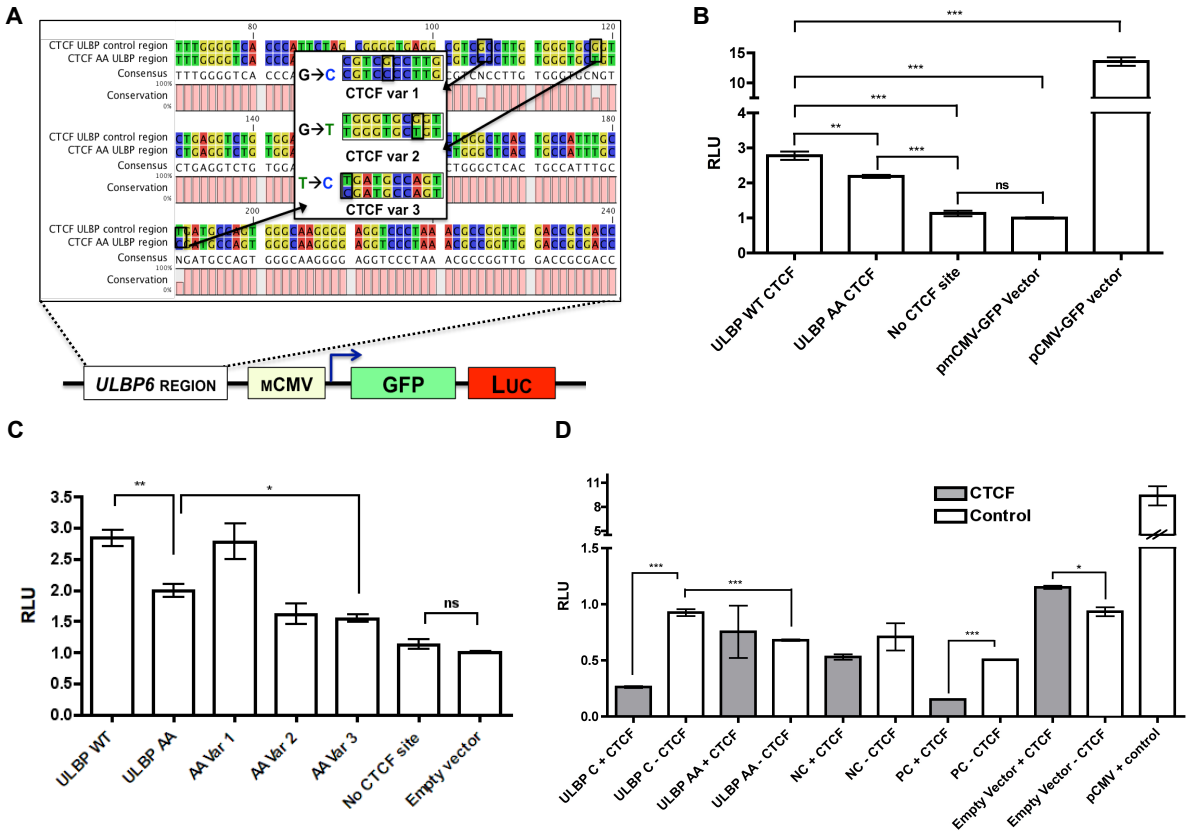


Figure 3. AA rare, non-coding variants disrupt *ULBP6* regulatory activity and abrogate CTCF binding
A. View of the *ULBP6* 250 bp region cloned into the pGreenFire reporter vector containing a minimal CMV promoter. **B.** Luciferase assay of the *ULBP6* control (wild-type) and AA rare variants sequences compared to a sequencing containing no CTCF site, and normalized to the empty vector. The pCMV-GFP vector containing the full CMV promoter was included as a positive control for the assay. **C.** Luciferase activity of the individual AA rare variants 1-3 compared to the control and AA sequences containing all variants (ULBP AA). **D.** The effects of CTCF overexpression on luciferase activity; NC = negative control sequence containing no CTCF site, PC = positive control region containing multiple CTCF sites.

Novel TFBS created by rare variants point to increased activation of *ULBP6* in AA

In our initial analysis of the rare non-coding variants at the *ULBP3/6* locus, we used RegulomeDB to test whether these variants are predicted to abolish TFBS motifs or other regulatory elements in the region. However, this analysis did not provide insight into whether the AA variants are predicted to create novel TFBS. We next performed an *in silico* analysis of the control and AA sequences using the software program TFBIND that searches for TFBS as well as TATA, GC, and CCAAT boxes. We identified two additional TFBS generated by two rare variants that were not found in the control sequence, specifically ELK1 and EGR3. ELK1 (ETS domain containing protein 1) is a transcriptional activator downstream of the ERK1/2 signaling pathway, reported to preferentially bind to the *IL10* risk allele in systemic lupus erythematosus (SLE) and become activated by ULBPs to mediate natural killer cell responses (SUTHERLAND *et al.* 2002; SAKURAI *et al.* 2013). EGR3 (Early growth response protein 3) is an NFAT target gene and inducer of NFKB during inflammation and antigen-induced proliferation of B and T cells (WIELAND *et al.* 2005; LI *et al.* 2012). Considering the known functions of these activating TFs in the immune system, we postulate that an increased binding affinity for ELK1 and EGR3 at the *ULBP6* regulatory region could also contribute to the sustained activation of *ULBP3/6* gene expression in AA.

CTCF-mediated long-range interactions at the *ULBP3/6* locus

Several immune-related genes are transcriptionally regulated in a spatiotemporally coordinated manner during lineage commitment, immune responses as well as other processes. These loci have been reported to engage in long-range interactions associated with changes in chromatin architecture to promote a more transcriptionally

active environment. This phenomenon has been demonstrated at the *HLA* locus, the Th2 locus, as well as several others, where the CTCF chromatin insulator/remodeling protein mediates chromatin-looping events to block or promote enhancer-promoter interactions, depending on the context (MAJUMDER *et al.* 2008; MAJUMDER AND BOSS 2010; GILLEN AND HARRIS 2011). Moreover, recent studies have demonstrated the presence of super enhancers (SEs) at well-characterized immune loci that serve as large clusters of lineage-specific transcriptional enhancers, where SEs are defined by p300 binding occupancy and H3K27Ac enrichment and CTCF demarcates these boundaries (HNISZ *et al.* 2013; FARH *et al.* 2015; VAHEDI *et al.* 2015). To determine whether the *ULBP3/6* locus possesses SE activity, we performed *in silico* analysis of the region using ENCODE data (Figure 4). We did not observe enrichment for p300 or H3K27Ac, however, we did identify clusters of CTCF binding in chromatin-accessible regions (DNAse 1 hypersensitive sites and ENCODE FAIRE-seq) (Figure 4).

In the context of alopecia areata and the loss of immune privilege in the hair follicle, the *ULBP* genes are strongly upregulated in the lesional hair follicle, raising the possibility that this locus participates in long-range interactions associated with gene repression in the healthy hair follicle. In our *in silico* analysis of the *ULBP3/6* locus, we observed several clustered sites of CTCF binding enrichment in multiple cell types, and observed overlap with ENCODE CTCF-mediated Chromatin Interaction with Paired-End Tags (ChIA-PET) experimental data that identified CTCF-mediated long-range interactions at this locus (Figure 4). Two of the rare variants lie directly within a paired-end tag (representing an interacting sequence) (rare variant cluster indicated by the blue vertical line), while the other paired-end tag is situated ~100 Kb downstream of

ULBP6 (Figure 4, gray bars).

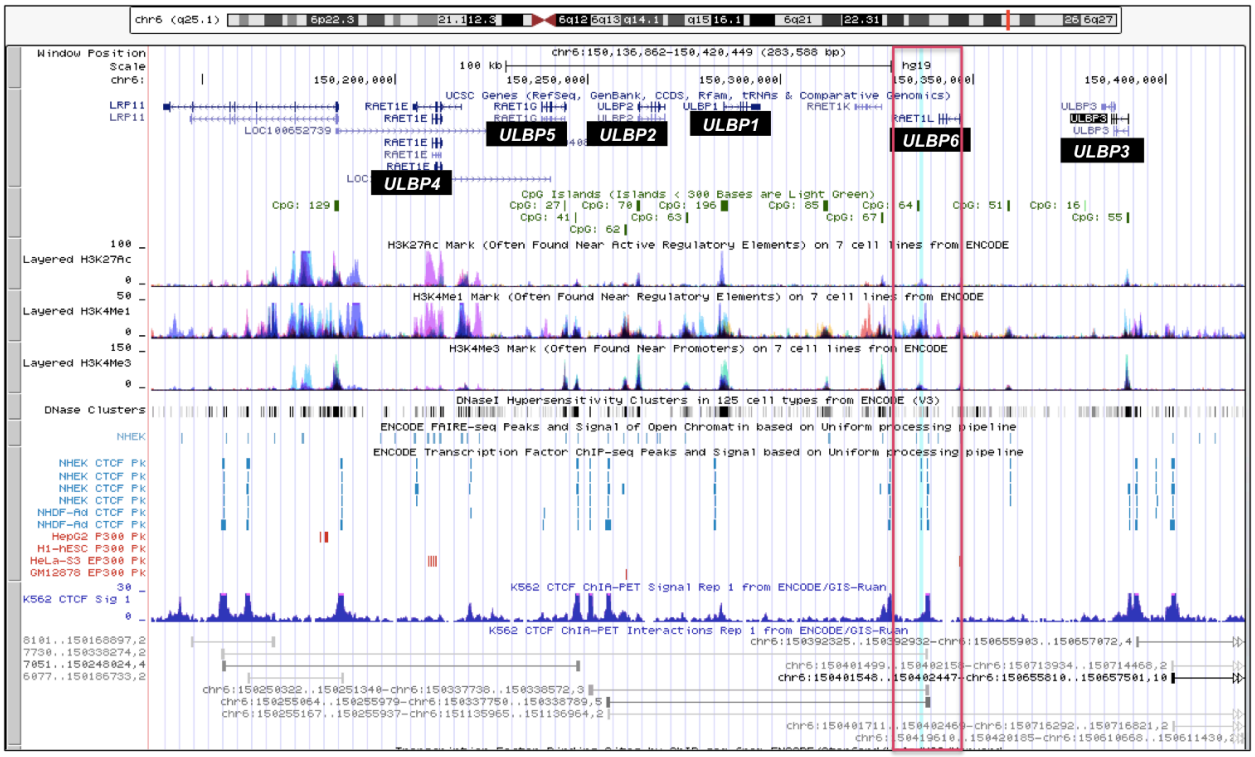


Figure 4. CTCF long-range chromatin interactions at the *ULBP6* locus. UCSC Genome Browser view of the *ULBP1-6* locus on human chromosome 6q25.1; *ULBP6* and *ULBP3* are located to the far right. Layered H3K27Ac (associated with active regulatory elements), H3K4me1 (associated with regulatory elements), and H3K4me3 (associated with promoter elements) are displayed. CTCF and p300 binding occupancies are displayed as blue and red bars, respectively, in normal human keratinocytes (NHEKs) and normal adult human dermal fibroblasts (NHDF-Ad). DNAse I hypersensitive sites and ENCODE FAIRE-Seq on NHEKs signifying open chromatin is shown. CTCF binding occupancy from ENCODE ChIP-seq in K562 cells is represented by the purple peaks, ChIA-PET interacting sequences are demonstrated by the long gray bars, as well as by the purple peaks.

This finding together, with those from our reporter assays (Figure 3), provides evidence to support the hypothesis that the rare, non-coding variants may disrupt regulatory activity and/or long-range interactions at the *ULBP3/6* locus in AA. Future functional studies utilizing long-range reporter assays and chromatin conformation capture (3C) with high throughput sequencing in human dermal fibroblasts will elucidate whether the *ULBP6* candidate region participates in long-range interactions, what the interacting sequences are, and whether the AA rare variants alter such interactions.

Discussion

For decades, the genetic architecture underlying alopecia areata has remained elusive, and our understanding of disease pathogenesis was limited to the knowledge of its autoimmune basis and the immunopathological findings in the skin characteristic of active disease. Recently, family-based studies at both the candidate and genome-wide levels have uncovered novel genes and loci associated with AA (MARTINEZ-MIR *et al.* 2007), but given its polygenic inheritance pattern, a more comprehensive approach was warranted to decode the complex genetic architecture.

The Christiano lab recently conducted the first GWAS for AA and identified eight regions of the genome that were significantly associated with the disease, where the highest signal outside the *HLA* emanated from the *ULBP3/6* locus, a region not previously associated with AA or any other autoimmune disorder. The *ULBP* genes encode ligands for the cytotoxic T cell receptor NKG2D whose signaling axis had previously been implicated in AA disease pathogenesis from candidate gene association studies (MICA/B ligands) (BARAHMANI *et al.* 2006). Moreover, the overrepresentation of allelic variants in other NKG2D ligands (MICA, MICB) had also been reported in other autoimmune diseases, including rheumatoid arthritis (RA), type 1 diabetes (T1D), celiac disease (CeD), and many of our GWAS regions of association were shared with these diseases (PETUKHOVA 2010).

Given the common dysregulation of NKG2D signaling but unique association of *ULBP3/6* with AA, the statistical associations at this locus were biologically validated by the presence of CD8+NKG2D+ T cells within the immune infiltrate and a strong upregulation of the ULBP3/6 ligands on lesional hair follicles (PETUKHOVA 2010).

Importantly, this finding was consistent with a previous report of the upregulation of another NKG2D ligand, MICA, on human lesional hair follicles, and collectively, these data placed the ULBP-NKG2D signaling axis squarely at the center of disease pathogenesis in AA. Importantly, our recent replication study in an independent cohort and meta-GWAS analysis not only validated but also increased the significance of the association signal for the *ULBP3/6* locus in AA (BETZ *et al.* 2015). Moreover, we recently demonstrated that CD8+NKG2D+ cells are both necessary and sufficient for disease pathogenesis in the C3H/HeJ mouse model for AA (XING *et al.* 2014). In this study, we determined the underlying pathogenic variants at the *ULBP3/6* locus driving the association with human AA.

Here, we performed targeted deep resequencing of the *ULBP3/6* locus in 124 AA patients and tested the hypothesis that rare, non-coding variants contribute to dysregulated transcription factor binding and regulatory activity of this locus. In the study of common diseases, the emerging picture of the genetic architecture is highlighting the contribution of rare variants to the “missing heritability,” therefore, we focused our studies on rare variants that are enriched on GWAS risk haplotypes. Importantly, our robust functional genomics approach enabled us to interrogate the functional consequences of these variants in the context of regulatory activity and transcription factor binding, and identified novel variants that reside within CTCF binding sites. We discovered that these variants alter the overall molecular properties of the *ULBP6* downstream sequence, since they decreased the reporter activity *in vitro*. Interestingly, we observed a stronger effect for variants 2 and 3 on decreasing reporter activity than the three variants combined, suggesting the presence of different *trans*-

binding factors and/or complexes, and that these variants are producing the strongest effect. We found that these variants created new TFBS *in silico* for stress-induced activating proteins with known roles in the immune system, and that the repressive effects of CTCF binding to this sequence are abrogated in the presence of the AA rare variants.

Functional genomics studies have been performed in other complex genetic disorders following deep resequencing of candidate regions to characterize the effects of susceptibility variants on regulatory activity and/or transcription factor binding (VERLAAN *et al.* 2009; HUTCHINSON *et al.* 2014; KAPOOR *et al.* 2014). Interestingly, a recent functional genomics study reported the association of non-coding variants that were identified from GWAS in several common autoimmune diseases (Celiac disease, IBD, Crohn's disease) and by SNP genotyping followed by genome-wide bisulfite sequencing, reported allele-specific methylation patterns associated with the variants (HUTCHINSON *et al.* 2014). Moreover, a functional genomics study at the *ZPBP2/GSDMB/ORMDL3* locus associated with asthma and several autoimmune disorders reported allele-specific chromatin remodeling, perturbed CTCF binding and altered nucleosome occupancy associated with non-coding variants (VERLAAN *et al.* 2009). Collectively, these studies highlight the functional and regulatory mechanisms by which non-coding variants may contribute to disease pathogenesis.

Several immune-related loci including the MHC cluster, the Th2 region (*IL4, IL5, IL13*), and *IL-1/IL-36/IL-37* gene cluster have been reported to engage in long-range chromatin interactions that govern the spatiotemporal regulation of these genes (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS

et al. 2013; SHARAF *et al.* 2014). Additionally, reorganization of the nuclear architecture and the formation of active chromatin hubs has been observed for some of these loci that engage in intra- and interchromosomal looping interactions upon differentiation or cytokine stimulation (JHUNJHUNWALA *et al.* 2009; MAJUMDER AND BOSS 2010). Many immune-related genes are monoallelically expressed and tightly coordinated regulation of gene expression in the immune system is key for many biological processes including V(D)J recombination in the generation of antigen diversity, as well as T cell lineage commitment (JHUNJHUNWALA *et al.* 2009).

Recent studies have demonstrated that lineage-specific super-enhancers (SEs), defined as large clusters of transcriptional enhancers for genes controlling cellular identity, are present at several well-characterized immune loci, and are detected by lineage-specific p300 binding occupancy and H3K27Ac enrichment (HNISZ *et al.* 2013; FARH *et al.* 2015; VAHEDI *et al.* 2015). A large-scale fine mapping study of GWAS variants conducted across many autoimmune diseases revealed that ~90% of causal variants are non-coding, and moreover, that ~60% of those variants reside within immune-cell enhancers and/or TFBS for master regulators of immune cell differentiation and activation (FARH *et al.* 2015). Most recently, T cell maps of SEs were generated and pointed to cytokine and cytokine receptor genes as the predominant source of SE architecture (VAHEDI *et al.* 2015). In the context of disease, T cell SEs were enriched at autoimmune disease-associated SNPs, where treatment of T cells with an inhibitor of the JAK/STAT pathway, tofacitinib, disproportionately affected the expression of SE-localized risk genes associated with rheumatoid arthritis (VAHEDI *et al.* 2015). Importantly, these studies identified key regulatory nodes in immune pathways and

have also defined a role for the insulator protein, CTCF in demarcating the boundaries of SEs across immune-related loci. Considering these findings, we utilized ENCODE data in both immune cells and keratinocytes and fibroblasts to examine the *ULBP3/6* locus for SE activity. While we observed clusters of CTCF binding enriched at sites of open chromatin (ENCODE FAIRE-seq and DNAse1 hypersensitive sites), we did not detect enrichment for p300 or H3K27Ac binding at these sites (Figure 4), suggesting that this region may not possess SE activity.

CTCF has several reported roles in the immune system, engaging in chromosome loops that bring immune enhancer elements in close proximity to promoters (facilitating gene expression), or physically creating barriers that separate genes from distal regulatory elements (promoting a repressive state) (PHILLIPS AND CORCES 2009). CTCF is a protein that homodimerizes with itself and heterodimerizes with the cohesin complex, components of the polycomb repressive complex, as well as many other factors (LI *et al.* 2008; RUBIO *et al.* 2008). The many functions of CTCF have been characterized in various cell types and systems, and several studies have demonstrated a clear inverse correlation between CTCF binding occupancy and the CpG methylation status, particularly at the *H19/IGF2* imprinting control region (ICR) (PHILLIPS AND CORCES 2009).

In the context of long-range interactions, we postulate that the *ULBP3/6* gene cluster is spatiotemporally controlled and chromosomally configured in a repressive state, such that these genes remain repressed in healthy cells. It remains to be determined whether these genes exhibit monoallelic expression and/or distinct chromatin configurations. Future studies to elucidate these questions will include

overlaying SNP genotyping and transcriptome-based data, as well as performing the Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE) assay that detects regulatory regions devoid of nucleosomes. In conjunction with ChIP-PCR for CTCF, this method can be used to determine allele-specific chromatin effects at the *ULBP6* region that may be perturbed in the presence of the AA rare, non-coding variants (SIMON *et al.* 2012).

To determine whether CTCF-mediated long-range interactions are present at the *ULBP6* locus, we analyzed ENCODE data from ChIA-PET experiments to capture CTCF-mediated interactions, and identified several interactions present at this locus in distinct cell types. Remarkably, we found that the AA rare non-coding variants lie precisely within three paired-end tags from CTCF ChIA-PET experiments, where the other paired-end tag lies ~100 Kb downstream of *ULBP6*, providing strong evidence for potential CTCF-mediated long-range interactions at this locus. Based on our findings, we suggest these interactions may be disrupted in the context of disease pathogenesis. Future experiments to test this hypothesis will include long-range reporter assays to determine whether the *ULBP6* candidate region can function as an insulator element in the presence of CTCF, ChIA-PET assays on healthy hair follicle fibroblasts to identify cell type-specific CTCF-mediated interacting sequences at this locus, and Hi-C (Chomosome Conformation Capture with high-throughput sequencing) to identify all intra- and interchromosomal interactions in both healthy and AA patient fibroblasts.

Although we provide evidence to support the hypothesis that AA variants disrupt CTCF-mediated long-range interactions at this locus, it is likely there are several additional binding proteins involved in these interactions. As previously mentioned,

CTCF has been reported to bind the cohesin complex, which is composed of four subunits (SSC1, SSC3, SMC1, SMC3) and is known to regulate the separation of sister chromatids during cell division (RUBIO *et al.* 2008). An *in silico* analysis of predicted TFBS motifs within the *ULBP6* region demonstrated SMC1 and SMC3 binding sites in close proximity to the CTCF sites. Moreover, CTCF is known to interact with components of the polycomb repressive complex 2 (LI *et al.* 2008) and we also identified predicted SUZ12 and EZH2 binding sites *in silico*, where one of the non-coding variants identified from our deep sequencing study lies within a predicted SUZ12 motif. Taken together, we postulate that these proteins together with CTCF orchestrate long-range interactions contributing to the spatiotemporal *ULBP3/6* gene expression.

In this study, we successfully utilized the functional genomics approach to uncover genetic susceptibility variants at the *ULBP3/6* locus, which we identified as a critical node of the NKG2D signaling axis in disease pathogenesis. We identified and characterized three rare and novel non-coding variants enriched on GWAS risk haplotypes, and found that they alter the molecular properties of the sequence, disrupt CTCF binding sites and create novel TFBS. In the context of the complex and dynamic regulation of immune-related genes and their association with SEs, we postulate that similar mechanisms employing long-range interactions control the *ULBP3/6* locus to ensuring proper spatiotemporal expression. Moreover, based on the findings from our functional experiments, we provide evidence that the rare and novel, non-coding variants identified in this study disrupt the regulatory activity and transcription factor binding properties in this region, further underscoring the biological significance of this locus in AA disease pathogenesis.

Materials and Methods

Ethics Statement

Informed consent was obtained from all subjects and approval for this study was provided by the Institutional Review Board of Columbia University in accordance with the Declaration of Helsinki Principles.

Patient information and materials

Patients from the previously published GWAS were ascertained through the National Alopecia Areata Registry (NAAR) and guidelines for enrollment as well as control cases used are described in (PETUKHOVA *et al.* 2010). GWAS samples were selected for targeted resequencing based on disease severity, family history of AA and enrichment of risk haplotypes. Genomic DNA samples isolated from whole blood were previously obtained from the GWAS cases, and subject DNA samples containing one or more rare variant enriched on GWAS haplotypes were selected for functional studies. Human dermal scalp fibroblasts were isolated from occipital scalp biopsies taken from healthy controls that are considered IRB-exempted (45 CFR Part 6) non-human materials. The human hTERT DP cell line used in ChIP assays was purchased through Applied Biological Materials (Cat. #T0501).

Targeted deep resequencing of the *ULBP3/6* locus

We designed a Raindance primer library to deep sequence 72 Kb of the *ULBP3/6* encompassing the 13 typed SNPs and 47 imputed SNPs from the GWAS. Targeted deep resequencing of the *ULBP3/6* locus was performed in 124 AA patients on the Roche 454 FLX platform, and the data were analyzed using the GATK pipeline (Broad Institute, Germany). Identification of rare variants was based on the novelty or rare

frequency of identified variants in public databases, and enrichment was defined as being prevalent in three or more AA cases. Functional annotations using ENCODE and RegulomeDB databases were performed to identify the non-coding rare variants with predicted regulatory function and/or in TFBS motifs.

Cloning and reporter assays

For all reporter assays, the pGF-mCMV vector (System Biosciences) harboring a minimal CMV promoter was used. The *ULBP6* region was amplified using a nested PCR: F1: 5'- TTCCCTCCTTCATCCTCGCGGGGAC-3'; R1: 5'- GCAATGAGGAAAGA GTCCTAATTC-3'; F2: 5- CTCTCAGTCTCCTGCTCTCTGCCTTGT-3'; R2: TGGGGTCGCGGTCCAACCGGCGTTTAG-3'. A region containing no CTCF sites was amplified using the following primers: F: 5'- GACACTGAGCCTTGTTATTCGTACATAC-3'; R: 5'- ATAAGCAGACCTAAGACCTCACAGAGC-3'. For the positive control of CTCF binding, an oligonucleotide sequence containing tandem CTCF sites was synthesized through Integrated DNA Technologies (IDT). The CREB sequence used for reporter assays is as follows: CATGGTGACGTCACTTGGTGACGTCACCTGGTGACGTC ACGTGGTGACGTCACATGGTGACGTCA. PMA (phorbol 12-myristate 13-acetate) was purchased through Sigma Aldrich.

Human dermal fibroblasts were nucleofected using the Amaxa Nucleofector and Basic Mammalian Fibroblasts reagents (Lonza) with 2ug of each DNA construct for the pmCMV-GFP vector, the CTCF-FLAG construct, as well as a control Renilla luciferase reporter. Luciferase activity was measured using the Promega Dual Luciferase assay kit on a 96-well plate reader. Firefly/Renilla luciferase ratios were calculated and samples were normalized to the empty vector.

Chromatin Immunoprecipitation (ChIP) and Immunoblotting

ChIP was performed using the EZ-ChIP kit reagents and protocol with the following modifications. Human dermal fibroblast cells isolated from scalp and the human hTERT DP cell line (Applied Biological Materials, cat. #T0501) were grown to confluency on 15 cm dishes (each containing 1.0-2.0 x 10^7 cells), cross-linked using formaldehyde (final concentration 1.1%), and then centrifuged for subsequent nuclei purification and sonication. Following cell lysis, cells were resuspended in sonication buffer, sonicated for 30-45 minutes (30" on/off cycles) using the Bioruptor bath sonicator (Diagenode). Fifty micro liters of nuclear lysate was saved for each sample as Input Chromatin (1/20 of the ChIP reaction). For antibody binding, 2ug was used for CTCF clone 2.24.3.10.1.4 from EMD Millipore (monoclonal antibody 1) and CTCF clone 1.2.1.5.4 from EMD Millipore (monoclonal antibody 2); 1ug antibody was used for RNA polymerase II clone CTD4H8 from EMD Millipore and normal mouse IgG (EMD Millipore cat. # 12-371B). Following the elution of protein/DNA complexes, crosslinks were reversed overnight at 65°C and protein was isolated prior to DNA purification for immunoblotting.

Immunoblotting for CTCF was performed using the protocol described in (DESTEFANO *et al.* 2014) to determine enrichment prior to proceeding with the PCR. As a positive control for CTCF detection, protein isolated from human dermal fibroblasts that were transfected with a FLAG-tagged form of CTCF (NM_006565 cDNA ORF clone, OriGene RC202416L2) was run on an SDS-PAGE gel and blotted for CTCF (monoclonal antibody 1) and FLAG (monoclonal antibody, OriGene clone 4Ca). The CTCF and FLAG antibodies were used at a concentration of 1:2000 and 1:000,

respectively, and the goat anti-mouse-HRP light chain-specific secondary antibody was used at a concentration of 1:2,500.

qPCR on ChIP DNA

Following ChIP for CTCF, qPCR was performed using standard methods and the protocol described in (DESTEFANO *et al.* 2014) to determine CTCF binding enrichment at the *ULBP6* region. Relative quantification was used (ddCt method) to normalize IP samples to Input Chromatin, where Ct values were adjusted for the Input Chromatin dilution by subtracting 4.322 (Log_2 20, dilution factor), and percent total input was calculated for all samples ($100*2^{-\Delta Ct}$). The primers used for qPCR analysis are listed below:

GAPDH
FOR: 5'-TACTAGCGGTTTTACGGGCG-3'
REV: 5'-TCGAACAGGAGGAGCAGAGAGCGA-3'
***H19* ICR**
FOR: 5'-CCCATCTTGCTGACCTCAC -3'
REV: 5'-AGACCTGGGACGTTTCTGTG-3'
***ULBP6* CTCF region**
FOR: 5'- GTGGACCAAGCTGACTAAGAATGA G-3'
REV: 5'- ACATGACTTTTCATGGTTTCGATCCTG-3'

Statistical analysis

A Student T test was used to calculate statistical significance ($p < 0.05$) in all reporter assays. Error bars in all figures represent SEM.

Supplementary Figure

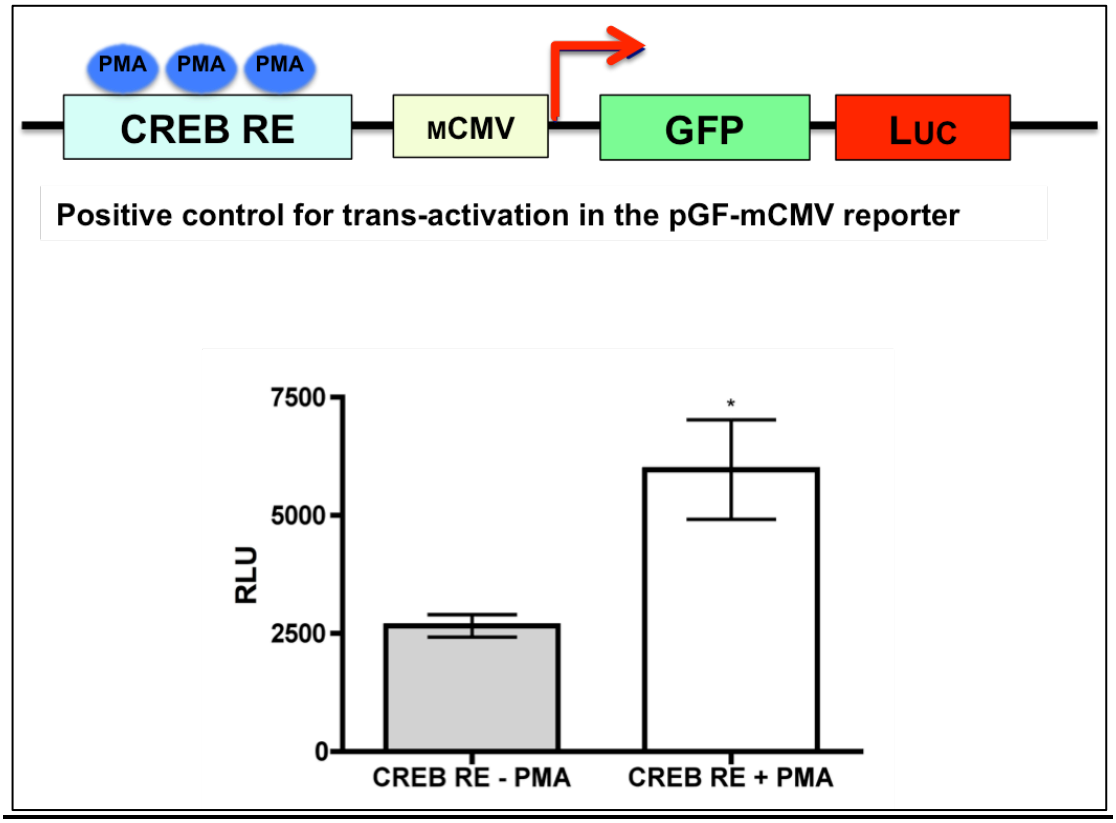


Figure S1. Scheme of the pGF-mCMV reporter construct and positive control for activity. The CREB sequence was cloned upstream of the mimimal CMV promoter and becomes activated upon the addition of PMA. Luciferase assay in dermal fibroblasts demonstrates CREB-driven reporter activity in the absence and presence of PMA. *p<0.05.

Chapter IV.5

Discussion

Chapter IV.5 Discussion

The search for causal genetic variants underlying complex, polygenic traits has proven to be a challenging task, due to the non-Mendelian patterns of inheritance and the contribution of many genetic and environmental factors. Therefore, GWAS became the popular approach to study these complex traits in populations rather than family-based cohorts, and this method has been performed for many well-defined traits, including diabetes, cancer, height, and several autoimmune disorders that display great heterogeneity across the population. While the GWAS has its limitations (detection threshold, cohort size required to see an effect, etc.), we now possess improved tools and technologies for next-generation sequencing that have already been utilized in several complex diseases and provided key insights into the distribution and contribution of variants (rare vs. common, coding vs. non-coding) to disease heritability (GOLDSTEIN 2009; ZHU *et al.* 2011).

Depending on the nature of the condition, GWAS studies conducted with larger cohorts may be fruitful in identifying novel susceptibility variants, but for most cases, this is not the case. Instead, GWAS performed for traits like height and type 2 diabetes have uncovered numerous variants associated with each trait, but that collectively only account for a small proportion of disease heritability. In light of these findings, the 'common-disease-common-variant' hypothesis posits that these polygenic disorders are composed of relatively frequent variants in the population that have low penetrances, or effect sizes, and thus require large cohorts. However, if we assume that for a given trait, nearly all the variants of large effect have been uncovered by GWAS, we would expect to see that upon increasing the cohort size, there is a "diminishing return" on effect size

distributions. In other words, it would take a lot of variants to account for the full heritability. Recent estimations of the number of common variants required to explain all the variance were calculated (GOLDSTEIN 2009), highlighting the implausibility of this model and its ability to identify the genetic component of a disorder.

On the opposite end of the spectrum, the 'common-disease-rare-variant' hypothesis posits that there are many rare variants with large effect sizes that account for a large percent of disease heritability. Based on the observations that common variants account for a small proportion of heritability for complex traits, low-frequency variants are more abundant in the human genome than common variants, and GWAS experiments cannot detect rare variants, the current view holds that the "missing heritability" can be accounted for by several rare variants with moderate effects, structural variants (i.e. CNVs), and genetic as well as gene-environment interactions (MANOLIO *et al.* 2009).

Importantly, next-generation sequencing approaches have afforded us the ability to detect the rare variants with larger effects and determine their contribution to disease pathogenesis. However, the most robust strategy to identify the missing heritability is to study both family-based and population-based cohorts for linkage analysis, whole-exome and genome sequencing, GWAS and subsequent replication(s)/meta-GWAS analysis, targeted deep resequencing of candidate regions, and CNV analysis to detect structural variants that would otherwise be missed. When prioritizing candidate regions for further investigation, it is of tremendous value to perform gene-expression based analyses (microarrays, RNA-seq) that can be correlated with the genetics, functional genomics studies to determine the consequences of intergenic variants on regulatory

function and transcription factor binding, as well as functional studies on mouse models that recapitulate genetic and/or pathological features of the human disease.

These approaches (discussed in Chapter IV.2-IV.4) enabled us to maximize our identification of susceptibility variants in alopecia areata. In the context of studying complex diseases, AA serves as a useful model and system, particularly in the context of autoimmune disorders, since there is strong evidence in support of a genetic basis and we have direct access to the pathological tissue as well as a mouse model that recapitulates the human pathology. Unlike other autoimmune disorders, the end organ (the hair follicle) is not destroyed, allowing us to perform detailed molecular and functional studies.

The immunopathological basis of AA is defined as the loss-of-immune privilege in the hair follicle leading to disfiguring hair loss, characterized by the infiltration of T lymphocytes at the base of the hair follicle. Interestingly, only growing, pigmented hair follicles are targeted, thus it is thought that AA is an antigen-driven disease, and that genetic dysregulation at both the level of the target (hair follicle) and effector cell (T cell) contributes to the pathology. Thus, we expect to identify genetic susceptibility variants in hair follicle-related genes and/or genes uniquely associated with AA, as has been observed from the GWAS and subsequent meta-GWAS analysis described in Chapter IV.2.

In the initial GWAS, eight regions of the genome were identified to be associated with AA, where several of these loci overlapped with type 1 diabetes, rheumatoid arthritis, and Celiac disease (PETUKHOVA *et al.* 2010). In these autoimmune conditions, the NKG2D pathway plays an important role in disease pathogenesis, thus the genetic

overlap at several immune-related loci further supported a role for the involvement of this pathway in AA pathogenesis. The Christiano lab uncovered two novel loci, *STX17* and *ULBP3/6* that are uniquely associated with AA. STX17 is a SNARE tethering protein involved in autophagy in melanocytes that is expressed in the hair follicle and has been shown to be associated with the grey hair phenotype in horses (SUNDSTROM *et al.* 2012; REISSMANN AND LUDWIG 2013; TEIXEIRA *et al.* 2013). The genetic association of *STX17* with human AA is biologically intriguing in the context of genes dysregulated in the end organ, as AA preferentially attacks the growing, pigmented hair follicle during active disease. In both the GWAS and meta-GWAS studies, the strongest association signal observed outside the *HLA* region came from the *ULBP3/6* locus, a novel region whose association is unique to AA. Given the known role of these genes in the NKG2D pathway and their upregulated expression in lesional hair follicles, we further investigated their role in AA and characterize the susceptibility variants driving the genetic association.

In Chapter IV.2-IV.4, we applied a robust strategy to define the genetic architecture and biological pathways that are disrupted in AA. Informed by the initial genetic studies, we identified the key mediators of cytotoxicity, defined their contribution to disease pathogenesis using functional studies in the C3H/HeJ mouse model, and then targeted them for therapeutic intervention in both mice and humans. In Chapter IV.2, we identified novel genes association with AA and replicated the previous findings in an independent cohort followed by a meta-GWAS analysis. We resolved the major histocompatibility complex (MHC) signal, and defined the independent risk variants located at the *HLA-DRA* and *HLA-DRB1* genes. Moreover, we uncovered two novel loci

associated with AA on chr2q13 and chr11q13.5, both of which have been previously associated with autoimmune disorders (BETZ *et al.* 2015). I characterized the novel BCL2L11 (BIM) gene on chr2q13 in association with AA. This gene encodes a pro-apoptotic factor that functions in immune cells as well as melanocytes, and is widely expressed in multiple cell types including epithelial and hair follicle cells (O'REILLY *et al.* 2000; LUO AND RUBINSZTEIN 2013). Moreover, since the regression stage of a normal cycling hair follicle, catagen, is characterized by apoptosis and cell death, and in AA, the hair follicle enters catagen prematurely due to autoimmune attack, BIM emerged as an interesting candidate gene to investigate in the contexts of both effector and target cell dysregulation.

In the work described in Chapter IV.2, I performed detailed molecular analyses of BIM and characterized its expression pattern in human immune cell populations, hair follicles, as well as within the mouse hair follicle. Consistent with its role in melanocyte apoptosis, BIM was expressed in keratinocytes and melanocytes in the human hair follicle bulb. Intriguingly, when I tested its expression pattern in mouse anagen, catagen, and telogen skin, BIM expression was restricted to catagen hair follicles, and in the context of disease pathogenesis using the C3H/HeJ mouse model, was more widespread and diffuse (BETZ *et al.* 2015). These data further support the hypothesis that BIM dysregulation may contribute to early entry into dystrophic catagen in AA hair follicles during active disease. To further implicate BIM in disease pathogenesis, the next steps following this study could include detailed functional analyses using mouse models, targeted deep resequencing of this locus to identify susceptibility variants in

human AA, and functional genomics studies to determine the consequences and effects of these variants in disease pathogenesis.

Considering our findings thus far and the implications of several important immune and hair follicle-related genes in AA disease pathogenesis, I next characterized the transcriptional profile of human and mouse lesional skin using an unbiased approach. The overarching goals of the work presented in Chapter IV.3 were to define the requirement for NKG2D+CD8+ T cells in disease pathogenesis and identify the key signaling molecules that drive cytotoxicity and can be targeted in the context of therapeutic treatment. In this study, I identified the genes with dysregulated expression using microarray analysis and transcriptome sequencing. Importantly, these studies identified three primary gene expression signatures that strongly implicated the NKG2D-ULBP signaling pathway and the role of cytokine genes and their receptors in immune cell activation during natural killer cell-mediated attack. The immunological findings and immunohistochemical analyses complemented the genetic studies in humans that revealed a strong association with these loci, and using bioinformatics tools to analyze gene expression data, we successfully identified key drivers of cytotoxicity, including IL15 and IL2, in interferon gamma-producing CD8+ T cells.

In the context of the bidirectional signaling and crosstalk events that occur between the target and effector cells to initiate destruction, it remains unknown in AA which comes first—the dysregulation at the end organ (the “chicken”) or within the T cells (the “egg”). Given the association of the NKG2D-ULBP signaling pathway in AA disease pathogenesis, and the presence of CD8+NKG2D+ T cells in the immune infiltrate, we tested the requirement for these cells in disease using the mouse model.

We first showed *ex vivo* in syngeneic dermal sheath cells that the T cells bearing the NKG2D receptor are in fact the cells mediating cytotoxicity during the autoimmune attack in AA. Using subcutaneous injections of total lymph node cells, NKG2D- cells, FAC sorted CD8+NKG2D+ T cells, and CD8+NKG2D- T cells, we demonstrated that CD8+NKG2D+ T cells are both necessary and sufficient to induce disease (XING *et al.* 2014).

Because these cells produce the cytokines and drivers of toxicity that are associated with the autoimmune response, we next targeted several type I cytokines (i.e. IL5 and IL2) in the mouse model using blocking antibodies. Although these antibodies alone did not reverse established disease, we learned that these cytokines signal through the JAK/STAT pathway, and thus, intervened downstream using FDA-approved small molecule inhibitors (Ruxolitinib and Tofacitinib) of JAKs 1-3. Importantly, this approach was successful in preventing and reversing the disease in the mouse model, and more recently, in treating human patients to fully restore hair growth (XING *et al.* 2014).

Collectively, these findings reinforced my interest in studying the human genetic variants in the NKG2D-ULBP pathway, thus, the goal of my work in Chapter IV.4 was to understand the mechanisms by which susceptibility variants at the *ULBP3/6* locus act to control the dysregulated expression of these genes in the human AA hair follicle. The GWAS revealed a strong association between the *ULBP3/6* locus and AA, where perhaps not surprisingly, the highest signal of association resided within an intergenic region of non-coding DNA. The emerging picture of the genetic architecture underlying complex disease highlights a role of rare and non-coding variants, and thus, I identified

and characterized the effects of the susceptibility variants at *ULBP3/6* associated with human AA. To accomplish this, we performed targeted deep resequencing of this locus, and identified rare and common variants that were enriched on GWAS risk haplotypes. In Chapter IV.4, we performed a detailed functional genomics study, characterizing the functions of three novel rare and non-coding variants identified from our targeted deep sequencing, and using publically available ENCODE data, determined that these variants resided within a presumptive regulatory region. Moreover, we crosschecked the variants against RegulomeDB and found that they were predicted to disrupt CTCF TFBS. *In vivo*, I used ChIP-PCR to detect endogenous CTCF binding at this locus, and *in vitro*, I found that CTCF exerts repressive effects on the *ULBP6* sequence, consistent with its known role as an insulator and repressor. The presence of the AA variants markedly reduced reporter activity and abolished CTCF-mediated repression, suggesting that these variants exhibit decreased affinity for CTCF binding at this locus.

In the AA patients harboring these rare variants, the associated risk haplotypes also possess common variants at this locus, whose effects I did not test in this study. Although the rare variants are likely to produce larger effects than common variants, it is possible that the common variants (in conjunction with the rare variants) produce an effect and contribute to the altered molecular properties of this sequence. Importantly, the data and findings from this study do not exclude the contribution of additional factors involved in the regulation of this locus. As discussed in Chapter IV.4, we postulated a role for other *trans* binding factors including the cohesin and PRC2 protein complexes in facilitating *ULBP3/6* long-range regulatory interactions.

Functional genomics studies performed in other autoimmune diseases have begun to uncover the mechanisms by which non-coding variants exert their effect and act at a given locus, and have implicated a role for heterozygous SNVs in allele-specific methylation and chromatin remodeling (discussed in Chapter IV.4) (HUTCHINSON *et al.* 2014) (VERLAAN *et al.* 2009). We uncovered a novel role for long-range interactions at the *ULBP3/6* locus, and CTCF as a candidate for mediating these interactions. This is further supported by the observation of CTCF ChIA-PET data from ENCODE experiments. It remains an open question as to whether the chromatin architecture of one *ULBP3/6* allele is distinct from the other, and whether this difference somehow influences how the homologous allele is regulated. Further functional studies examining the chromosomal architecture of this region will elucidate this phenomenon and are discussed in Chapter IV.4.

In the context of an immune response, several immune-related loci engage in CTCF-mediated long-range regulatory interactions, some of which have been shown to possess super-enhancers whose boundaries are demarcated by CTCF binding (HNISZ *et al.* 2013; FARH *et al.* 2015; VAHEDI *et al.* 2015). Although the *ULBP* gene cluster was not identified in previous studies that reported these super-enhancers, and this region is not enriched for p300 loading (a characteristic feature of SEs) in keratinocytes or fibroblasts (data not shown), we identified large peaks of H3K27Ac (mark of active regulatory elements) demarcated by CTCF ChIP-seq binding, as well as clusters of DNAse I hypersensitivity sites. Therefore, it remains possible that this region possesses SE-like structures that contribute to the upregulation of these ligands during an immune response or targeted autoimmune attack.

The *HLA* genes and regulatory mechanisms controlling their spatiotemporal expression have been well characterized and are known to involve complex intrachromosomal interactions resulting in conformational changes in chromatin architecture (MAJUMDER *et al.* 2006; MAJUMDER *et al.* 2008; MAJUMDER AND BOSS 2011). Considering what is known about the genomic mechanisms regulating these and several other immune-related genes (JHUNJHUNWALA *et al.* 2009; SEKIMATA *et al.* 2009; MAJUMDER AND BOSS 2010; WILLIAMS *et al.* 2013; SHARAF *et al.* 2014), I postulate that the *ULBP3/6* locus may be under similar control and in a tightly coordinated fashion to ensure that these genes remain off in healthy cells (i.e. the hair follicle). Interestingly, the *ULBP* gene cluster resides on the opposite arm of chromosome 6 relative to the *HLA* locus, raising the question of whether conformational changes in chromatin structure and movement at one arm constrain or in some way affect the overall configuration of the opposite arm. In other terms, could the conformational changes at the *HLA* locus influence the configuration of the *ULBP* locus? Time-lapsed *in vivo* imaging of fluorescently tagged chromosomes, FISH analysis of these regions, and chromosome conformation capture-based methods (discussed in Chapter IV.4) following cytokine stimulation may provide some insight into this phenomenon.

Considering the nature of immune privilege in the hair follicle that is in part characterized by the active repression of MHC and other immune-related genes (i.e. *ULBPs*), it invites the question of whether these genes are regulated by similar mechanisms, and perhaps by the same factors. Based on the collective findings from our comprehensive study interrogating the genetic, molecular and immunopathological bases of AA, I postulate that long-range interactions at the *ULBP3/6* locus may serve as

a genomic mechanism regulating the spatiotemporal expression of these genes (maintaining active repression in healthy tissues), and that in the context of disease, non-coding susceptibility variants identified in AA patients abrogate the regulatory and transcription factor binding properties, resulting in perturbed intrachromosomal interactions.

In conclusion, the studies described in Chapter IV of this thesis and the approaches we used in this work demonstrate innovative strategies to dissect the genetic architecture underlying complex genetic disease. The pursuit of this study effectively began over 15 years ago, where after completing the first GWAS for AA, the Christiano lab was immediately faced with a daunting list of associated SNPs and genomic regions whose biological relevance and contribution to disease pathogenesis was unknown. Combining the GWAS data with functional studies in the mouse model, next-generation sequencing methods and functional genomics tools provided us with mechanistic insight into the genetics and genomics underlying the pathology of AA.

Chapter V

General Discussion

Chapter V. General Discussion

The genetic underpinnings of human hair growth are complex, invoking several mechanisms to regulate gene expression. As such, hair follicle morphogenesis and growth are tightly regulated processes, where perturbations lead to abnormalities in the size, pattern, structure, and number of hair follicles. Thus, we can study genetic conditions affecting hair growth as models to interrogate the molecular basis of the pathology, and obtain novel insight into mechanisms and pathways required for normal hair growth. The hair follicle is a miniorgan with a highly complex and dynamic structure, providing an excellent system to study genetics because of its readily accessible nature and continuous cycling through the lifetime of the animal. Moreover, the key developmental pathways and diverse range of cell types in the hair follicle are well characterized, and the spatiotemporal regulation of gene expression has been widely studied.

When studying genetic conditions affecting hair growth, the approaches and tools we use depend largely on the prevalence of the trait in the general population and the observed inheritance pattern. We can readily predict the transmission of traits that are inherited in a Mendelian fashion and use the classic genetic approach to identify the genes involved. However, there are the rare, extreme traits we don't see too often in the general population (such as excessive hair overgrowth all over the body), and these sporadic conditions have historically been difficult to study using positional and functional cloning approaches. On the opposite end of the spectrum, many genes plus environmental factors control common, complex polygenic traits like male-pattern baldness and alopecia areata, and so these traits have been difficult to study because

they do not follow a classic Mendelian inheritance pattern.

The completion of the human genome sequence in 2003 has revolutionized the field of genetics and genomics, providing a transformative tool in medicine that has greatly facilitated the identification of disease-associated genes. With the advent of GWAS studies a little over a decade ago, large-scale population-based studies to identify regions of association became the preferred choice in tackling the complex genetic architecture of non-Mendelian diseases. Moreover, next-generation sequencing methods and technologies have become standard practice over the past eight years, far surpassing the prediction of a two-fold improvement in the speed and cost of genome sequencing per year (MOORE 1965; MARDIS 2008). These novel and sophisticated approaches have greatly facilitated the identification of pathogenic variants in monogenic, Mendelian disorders and complex genetic diseases alike. In the former, these tools have enabled us to perform whole-genome/exome and targeted deep sequencing in the extreme cases of rare, sporadic traits whose genetic lesions remained elusive. And in the latter, these tools have enabled the capture of all genetic variation residing within GWAS regions of association. By targeted deep resequencing, we can now identify the rare variants that were previously undetected by GWAS, and perform further functional studies to determine their contribution to disease pathogenesis.

The goal of my thesis research was to identify and characterize the genetic mechanisms controlling human hair growth. To accomplish this, I studied three inherited conditions affecting human hair growth, from the rarest syndromes of excessive hair overgrowth, to a very common autoimmune disease of hair loss, alopecia areata, and

the approaches I used to detect the underlying variants in these conditions were contingent upon the inheritance patterns and overall prevalence within the population. In the first part of my thesis, I identified structural and genomic effects on human hair growth in a rare, sporadic Mendelian disorder using a combination of CNV analysis and next-generation sequencing methods (Chapter II). In the second part, I characterized single-gene effects on human hair growth in a rare, familial condition using whole-exome sequencing (Chapter III). Lastly, I performed a functional analysis of rare, non-coding variants in a complex polygenic autoimmune disease utilizing functional genomics (Chapter IV). It is important to note that throughout the work described in this thesis, our understanding of the genetic mechanisms perturbed in human disease were not only facilitated by, but also dependent upon, the availability of a comprehensive, detailed map of the human genome, as well as several novel and sophisticated technologies in next-generation sequencing.

Rare Structural Variants Controlling Human Hair Growth

Rare Mendelian diseases are defined as affecting fewer than 200,000 people in the population (in the United States) and the number of rare diseases recorded in OMIM is currently estimated to be between 6,000-7,000 (BOYCOTT *et al.* 2013). The rate of an intragenic mutation per gene (for an average gene length of 1,500 bp coding sequence), copy, and generation is approximately 5.6×10^{-6} for silent mutations, 1.28×10^{-7} for missense mutations, and 1.7×10^{-6} for disruptive mutations (ZUK *et al.* 2014). In contrast, the estimated incidence of inherited hypertrichosis syndromes is far lower, highlighting the presence of unusual genetic mechanisms underlying this rare sporadic

condition. Consistent with this observation, the Christiano lab and other groups have previously identified *de novo*, complex structural variants in the autosomal dominant and sporadic forms of hypertrichosis whose genetic basis have not previously been identified. Thus, it has been suggested that these syndromes are genomic disorders, characterized by structural and numerical chromosome anomalies (ZHU *et al.* 2011a).

In Chapter II, I identified the effects of structural and genomic defects on human hair growth in a large Mexican family with X-linked hypertrichosis in which the Christiano lab had previously reported linkage to chromosome Xq24-27 (FIGUERA *et al.* 1995; TADIN-STRAPPS *et al.* 2003). Using a combination of cytogenetic methods and next-generation sequencing technologies to detect structural variants, we identified a large interchromosomal insertion at a palindromic site that completed cosegregated with the disease phenotype (DESTEFANO *et al.* 2013), and was consistent with findings in two unrelated XLH families that possess interchromosomal insertions at the same palindrome site on Xq27.1 (ZHU *et al.* 2011a). However, the sequence content and sizes of the insertions are distinct across the three families. Given the rarity of these syndromes and the severity of the phenotypes, the probability of identifying three independent events characterized by such unusual genetic mechanisms is extremely small. Thus, this observation raised the question, how do these interchromosomal insertions exert their pathogenic effects and cause excessive hair overgrowth?

Several plausible mechanisms by which the insertions can disrupt hair growth include the disruption of a gene in the region, the introduction of novel regulatory elements or dosage-sensitive genes, separation of the nearby genes from their regulatory elements and altering expression levels, and disruption of the chromosomal

architecture and inter/intrachromosomal interactions. I deduced that the latter two mechanisms were most likely given that the insertions reside within a gene desert, do not contain full open reading frames for protein-coding genes or any genes involved in hair follicle development, and the size and content of the insertions is different among the three families.

Therefore, I tested the hypothesis that the interchromosomal insertion we identified altered the expression levels of the genes in the surrounding region, and that the candidate gene(s) would be expressed in the human hair follicle. Using qRT-PCR and RNA-seq, I determined the expression levels of the genes in the surrounding region and observed a significant and selective decrease in *FGF13*, a gene not previously implicated in human hair growth. I characterized its expression pattern in both human and mouse hair follicles, as well as in the context of XLH pathology (DESTEFANO *et al.* 2013). My interest in *FGF13* as a potential candidate emerged from the observation that FGFs play very important roles in hair follicle morphogenesis and cycling, one key example being *FGF5,* whose mutations increase the duration of anagen and cause the *angora* excessive hair overgrowth phenotype in mice and many other organisms (HEBERT *et al.* 1994). However, it is very likely that the interchromosomal insertion mediates pathogenic effects through a variety of mechanisms, including the disruption of important inter- and intrachromosomal interactions, and future studies examining genome-wide chromosomal interactions (i.e. using Hi-C, discussed in Chapter II) will provide insight into perturbed or ectopic interactions in XLH.

The overall incidence of X-linked disorders in the population is estimated to be between 1:10,000-1:30,000, thus, if the interchromosomal insertions identified in the

three X-linked hypertrichosis families exerted their pathogenic effects on a single gene (such as *FGF13*), one would expect to find intragenic mutations within the population at the same frequency as the predicted mutation rate for that gene, concurrent with the excessive hair overgrowth phenotype. Moreover, the affected family members of two of the X-linked hypertrichosis families possess additional congenital anomalies, including scoliosis and spina bifida (Chinese family, 125 Kb insertion) (ZHU *et al.* 2011a), and deafness, dental, as well as palate anomalies (Mexican family from our study, 389 Kb insertion), suggesting the dysregulation of several genes. Given the rarity of these syndromes and the identification of *de novo* structural variants, I postulate that the interchromosomal insertions exert several downstream effects and disrupt the spatiotemporal regulation of genes required for hair follicle patterning and morphogenesis.

The Human Hair Follicle and Developmental Patterning

During hair follicle morphogenesis in humans, the lanugo hairs, which are uniformly long, fine and unpigmented, and first appear on the scalp and forehead. Prior to the time when these hairs are shed *in utero*, the developing fetus appears to be covered in hair, where the same follicles on the scalp are also present on the forehead and face—a feature not observed in adult humans. These hair follicles are then shed in a synchronous manner and replaced by vellus hairs on the body and terminal hairs on the scalp, where every subsequent hair cycle is asynchronous. In the context of hypertrichosis, the excessive hair overgrowth phenotype is present at birth (congenital), and the hairs that cover the body are typically of the terminal type. Importantly, this

observation suggests that the genetic mechanisms underlying these syndromes perturb early events during hair follicle patterning and morphogenesis. I postulate that the signaling pathways that regulate the formation of a vellus versus a terminal hair and the boundaries demarcating these regions are altered by the genetic defect(s), and lead to a transformation in the type of hair produced. Although the precise mechanism(s) by which genetic mutations cause an excessive hair overgrowth phenotype in hypertrichosis are not clear, I postulate a role for the classical developmental signaling pathways (such as Wnt, Shh, and Bmp) whose roles in patterning have been well-characterized and widely studied in model organisms.

Given the well-defined and conserved role of the Wnt signaling pathway in developmental patterning, it has been suggested that the master regulators of human hair follicle patterning are Wnt-related genes, since gain-of-function mutations lead to excessive hair overgrowth (overexpression of stabilized β-catenin in the epithelium induces extra hair follicles in mice), and loss-of-function mutations block hair follicle induction (β-catenin ablation or overexpression of the Wnt inhibitor, *Dkk1*, in mice abrogates hair follicle morphogenesis) (HELD 2010) (ANDL *et al.* 2002). Interestingly, the key regulator of moustache and eyebrow growth in dogs is the *R-spondin 2 (RSPO2)* gene (CADIEU *et al.* 2009), a regulator of Wnt signaling. In humans, mutations in the *WNT10A* gene cause sparse hair across the body, scalp, eyebrows and eyelashes (ADAIMY *et al.* 2007). While there is strong evidence for the involvement of Wnt signaling in hair follicle patterning, genetic studies in humans and mice have demonstrated that there are many additional genes required for hair follicle patterning and morphogenesis that do not reside within the Wnt pathway. I postulate that the underlying genetic

defect(s) in human hypertrichosis syndromes disrupts several important signaling pathways required for proper hair follicle patterning, leading to an excessive hair overgrowth phenotype.

From flies to mice, studies in model organisms have elucidated the key genes and signaling pathways whose disruption or ectopic expression causes severe patterning defects and homeotic transformations. Genetic studies performed decades ago have demonstrated that many well-conserved key developmental processes and biological traits are controlled by discrete loci that encompass the necessary genes and regulatory elements for proper spatiotemporal control of gene expression. In the fly, the Achaete-Scute Complex (AS-C) is a well-studied locus that controls bristle patterning, and the expression of the *achaete* and *scute* genes is spatiotemporally regulated by nearby *cis*-enhancers (GOMEZ-SKARMETA *et al.* 1995). Similarly, the well-conserved *Hox* genes in metazoans that control anterior-posterior patterning during embryogenesis are clustered into paralogous groups, exhibiting spatiotemporal regulation and the unique property of colinearity, where their chromosomal arrangement correlates with their expression domains and the order of the structures they form along the body (MALLO *et al.* 2010) (DUBOULE 1998).

In the context of the human hair follicle, it has been speculated that the genes controlling hair patterning and the boundaries separating different hair types might be spatially arranged or clustered at a genomic locus that operates as a “Hair Headquarters,” similar to the AS-C locus in the fly (HELD 2010). In humans, this phenomenon is observed in the control of epidermal differentiation, where the genes comprising the Epidermal Differentiation Complex (EDC) on chromosome 1 are

coordinated in a spatiotemporal manner (OH *et al.* 2014). Similarly, the type I and type II keratin intermediate filament genes are arranged in a cluster on human chromosomes 17q12-21 and 12q13, respectively, whose coordinated activation is required for keratinization within the differentiating cells of the hair follicle (POWELL *et al.* 1991; POWELL AND ROGERS 1997; DUNN *et al.* 1998). While the Christiano lab and other groups have identified several genes that control hair follicle growth (SHIMOMURA AND CHRISTIANO 2010), these genes are not linked or clustered within a single locus. Given the observation that hair follicle morphogenesis and cycling are tightly coordinated processes controlled by well-defined signaling pathways, I postulate that several hair patterning-related genes are co-regulated in a spatiotemporal manner, perhaps facilitated by the same transcription factors, regulatory elements and/or long-range chromosomal interactions.

Evolutionary Atavisms and Hair Growth

Inherited hypertrichosis syndromes have long fascinated evolutionary geneticists because they are considered recurrences of an ancestral phenotype, or atavisms, which are quite extraordinary. Despite their rarity, they have been observed in several species and included the presence of a tail on a newborn human, extra toes in extant horses, hind limbs in whales, and a reptile-like coronary circulation and myocardial architecture in humans (HALL 1984; CANTU AND RUIZ 1985; HALL 1995; HALL 2010). The occurrence of these ancient traits suggests that the genetic potential to create these structures has been retained throughout evolution, and further highlights the presence of unusual mechanisms that are invoked to reactivate these genes.

Humans have lost a significant amount of their hair compared to non-human primates; thus, it has been postulated that the genes promoting a fuller coat were silenced during evolution (i.e. not under purifying selection). To date, no hair-related genes have been reported among the loci within the selective sweeps (regions positively selected for) of the human genome. However, there are several reports on the existence on pseudogenes in the type I keratin gene family locus on chromosome 17q21.2 that have not been identified in rodents (HESSE *et al.* 2004; ROGERS *et al.* 2004), but it remains unclear whether they are present in non-human primates who are evolutionarily closer to humans. Importantly, studying the inherited forms of hypertrichosis has provided us with important clues as to the genes and loci that may have contributed to a fuller coat of hair in non-human primates.

To gain insight into atavistic genes at the Xq27.1 locus in XLH, I performed comparative genomic analyses of DNA sequences, but this did not reveal the presence of non-human primate-specific genes or human-specific pseudogenes. It remains to be tested whether the presence of the insertion alters the conformational structure of the X chromosome and causes ectopic non-homologous chromosomal interactions with sequences that encode important patterning genes during hair follicle morphogenesis. One interesting candidate to examine would be the *Glypican-3 (GPC3)* gene at Xq26, residing nearly 6 Mb away from the interchromosomal insertion, that encodes a cell surface heparan sulfate proteoglycan and regulates the *IGF* and *WNT* signaling pathways. *Gpc3* loss-of-function mutations in the mouse leads to cellular and tissue overgrowth via inhibition of the non-canonical Wnt pathway and by increasing the signaling activity of the canonical pathway (CHIAO *et al.* 2002) (SONG *et al.* 2005). *GPC3*

mutations in humans cause Simpson-Golabi-Behmel syndrome, characterized by tissue overgrowth and several congenital defects and demonstrating phenotypic similarity to Beckwith-Wiedemann syndrome (PILIA *et al.* 1996; LI *et al.* 2001; JAKUBOVIC AND JOTHY 2007). Simpson-Golabi-Behmel patients do not display an observable hair defect, but it would be interesting to determine whether perturbations in *GPC3* spatiotemporal expression (rather than function) during morphogenesis can alter the Wnt signaling pathway, and perhaps lead to defects in hair follicle patterning.

Identifying Single-Gene Effects on Human Hair Growth

The mutations underlying rare, sporadic conditions affecting hair growth are often not intragenic, but rather, variants in chromosome number or structure that are complex and likely produce several downstream effects. Thus, it is often difficult to identify the precise mechanism(s) by which these variants act to produce the phenotype. The study of rare, Mendelian disorders affecting hair growth overcomes this limitation, since the genetic architecture is mostly composed of intragenic variants (exonic, splice site) with large effect sizes. By examining families with conditions affecting hair growth, the Christiano lab and others have successfully identified the pathogenic mutations that cosegregated with the disease phenotype providing novel insight into the genes and mechanisms involved.

In families with a rare autosomal dominant condition of hair loss called hereditary hypotrichosis simplex (HHS, characterized by hair follicle miniaturization), the Christiano lab performed genetic linkage analysis to map the disease locus, and identified a point mutation in the *APCDD1* gene encoding a membrane-bound glycoprotein. Remarkably,

the same point mutation was identified in three distinct families (two Pakistani, one Italian), providing strong evidence for the involvement of this gene in HHS pathogenesis (SHIMOMURA *et al.* 2010). Moreover, high *APCDD1* expression was observed in both mesenchymal and epithelial compartments of the hair follicle, consistent with a role for this gene in hair follicle miniaturization, and APCDD1 was identified as a novel inhibitor of the Wnt signaling pathway. These findings thus provided key insight into the underlying mechanisms and pathways involved in hair follicle miniaturization, a process by which thick, terminal hairs are progressively transformed into fine, vellus hairs. Collectively, our lab has studied inherited conditions of both hair loss and excessive hair growth, which afforded us the ability to identify the mechanisms controlling human hair growth.

In Chapter III of my thesis, I characterized the genetic defect underlying a rare, familial condition affecting hair growth by studying a small consanguineous family with congenital generalized hypertrichosis terminalis (CGHT). Using whole-exome sequencing, we identified a novel splice site mutation in the ABC transporter gene, *ABCA5* that cosegregated with the disease phenotype in a homozygous recessive manner, and subsequently demonstrated that the mutation leads to the complete loss-of-function (DESTEFANO *et al.* 2014). Importantly, the defining aspect of our filtering strategy that pointed to *ABCA5* as the candidate gene was its genomic position, where CNVs in chromosome 17q23-24 encompassing *ABCA5* were previously associated with CGHT (SUN *et al.* 2009; FANTAUZZO *et al.* 2012). Using several methods of detection, I found that *ABCA5* is highly expressed in both epithelial and mesenchymal compartments of the human hair follicle, and that its expression is conserved within

mouse anagen hair follicles. Moreover, *ABCA5* expression and localization was markedly reduced in CGHT patient hair follicles, the site of pathology for the phenotype, providing strong evidence for a complete loss-of-function (DESTEFANO *et al.* 2014).

Upon the discovery of novel genes associated with a human disease phenotype like excessive hair overgrowth, the next question in understanding the functional consequences is, what are the mechanisms by which these genes act? To answer this, I tested ABCA5 function in the context of lysosome function and cholesterol transport, since previous reports in *Abca5*$^{-/-}$ mice have demonstrated a role for this transporter in these processes (KUBO *et al.* 2005; YE *et al.* 2010). Remarkably, we found that *ABCA5* loss-of-function leads to reduced lysosome function, an accumulation of autophagosomes, autophagosomal cargos and increased endolysosomal cholesterol in CGHT keratinocytes (DESTEFANO *et al.* 2014). These data highlight a role for ABCA5 in controlling the fate of lipoprotein-derived cholesterol in humans, and are consistent with observations in mouse models.

In conjunction with my studies in Chapter III defining the genetic basis of autosomal recessive CGHT, I also analyzed an unrelated sporadic CGHT case from Mexico whose excessive hair overgrowth phenotype closely resembled that of the autosomal recessive CGHT case. Using cytogenetic methods and CNV analysis to detect structural variants, we identified a *de novo* t3;17 translocation that led to a cryptic 1.3 Mb deletion of chr17q24.2-24.3 encompassing *ABCA5*. Thus, I postulated that *ABCA5* expression and/or function is compromised in the sporadic CGHT case, and using detailed molecular and functional analyses, showed that *ABCA5* expression and protein localization was significantly reduced in patient hair follicles (DESTEFANO *et al.*

2014). Collectively, the findings from our study point to a prominent and evolutionarily conserved role for this gene in regulating hair growth.

Based on my findings together with previous reports in the literature, I postulate that ABCA5 possesses multiple functions in the hair follicle and plays an integral role in both lipid transport and hormone signaling pathways. Similar to the expression of the *P2RY5* and *LIPH* genes that regulate lipid transport in the hair follicle, *ABCA5* is strongly expressed in the IRS of the human and mouse hair follicle. Moreover, its dual expression in epithelial and mesenchymal lineages of the hair follicle (similar to that of *APCDD1*) suggest a role for this transporter in regulating the transformation of vellus to terminal hairs (where *APCDD1* loss-of-function causes hair follicle miniaturization).

Studies examining ABC transporters in human disease have shed light onto the significance of the allocrite as the key determinant of the phenotype. It remains an open question as to what specific molecule(s) are being transported by ABCA5 in the human hair follicle, as there is collective evidence for the involvement of both cholesterol and thyroid related hormones as substrates based on functional studies in mice. It may be that ABCA5 functions differently in the skin and hair follicle, compared to its functions in the heart and thyroid gland, where the latter two organs progressively deteriorate and eventually collapse in the knockout mouse. In contrast, the hair follicles appear to be intact and based on photos from the publication, the mice clearly possess a full coat of hair, though it currently remains unknown whether *Abca5* deficient mice possess an excessive hair overgrowth phenotype because this was not directly tested.

In the context of CGHT, we examined ABCA5 expression and function only within the skin and hair follicle, as we did not have access to other cell types. Thus, it

would be interesting to determine whether tissues that do not express ABCA5 possess any pathological defect, as well as whether the subcellular defects in the endolysosomal pathway observed in CGHT keratinocytes are also present within other cell types. In light of our observations, I postulate that the accumulation and sequestration of free cholesterol inside lysosomes may exert consequences at the subcellular level, perhaps by altering the physical and/or signaling properties of cytosolic and membrane-bound proteins that require cholesterol as a covalent modification to function. It is interesting to speculate that the activity of such proteins functioning in hair follicle morphogenesis may be altered due to scarce levels of free cholesterol at the cell periphery or within the cytosol. Further functional studies would be required to directly test this hypothesis.

Histological analysis of CGHT patient hair follicles from the forearm region revealed large pigmented hairs in anagen, where control hair follicles isolated from this same site were exclusively in catagen and telogen (DESTEFANO *et al.* 2014). Importantly, quantification of the lengths of the hair shafts revealed a significant increase in CGHT hair length, suggesting that CGHT hair follicles may possess an increased duration of anagen. Interestingly, I noticed that the ORS cells of the CGHT hair follicles appeared enlarged compared to controls—an effect I also observed in cultured CGHT keratinocytes (data not shown), and by hematoxylin and eosin staining, the cytoplasmic region of the outer most cellular layer of the ORS is white (instead of pink). This effect is consistent with the well-characterized observation that growing ORS cells and the stem cells in the bulge region are rich in glycogen particles (ORR *et al.* 1977; VSEVOLODOV *et al.* 1983; IMCKE *et al.* 1987; YAMAUCHI AND KUROSAKA 2010). Interestingly, humans with glycogen storage disease type II (Pompe disease, autosomal

recessive) possess an accumulation of glycogen particles in the lysosomes resulting from mutations in the lysosomal acid alpha-glucosidase enzyme (KO *et al.* 1999; FERNANDEZ-HOJAS *et al.* 2002; MANGANELLI AND RUGGIERO 2013). Since ABCA5 is essential for the integrity of lysosomes, I postulate that this transporter may participate in the efflux of glycogen in hair follicle keratinocytes, and functional studies in the CGHT patient cells as well as the *Abca5*$^{-/-}$ mouse model will provide further insights.

In the autosomal recessive CGHT case, the family we studied is consanguineous, and thus, the proband is homozygous recessive for several variants in regions of the genome that are identical-by-descent (IBD). In some autosomal recessive rare Mendelian conditions resulting from consanguinity, whole-exome sequencing studies have revealed the involvement of more than one pathogenic variant, especially in cases where an atypical phenotype is produced that resembles two separate monogenic disorders (i.e. patients with oculocutaneous albinism and neutropenia) (CULLINANE *et al.* 2011). Thus, it remains an open question as to whether the phenotype of excessive hair overgrowth, gingival hyperplasia and epilepsy observed in the autosomal recessive CGHT case is caused by more than one pathogenic variant and/or genetic interactions. An interesting candidate to test is the *DKGZ* gene that encodes a diacyglycerol kinase important for lipid signaling in the brain and in T cells (MEHTA *et al.* 2002; LUO *et al.* 2004; GHARBI *et al.* 2011; RINCON *et al.* 2012) (SCHMIDT *et al.* 2013). Based on studies in knockout mouse models, it is likely that the *DKGZ* substitution mutation in autosomal recessive CGHT is causal to the patient's seizures.

Interestingly, DKGZ phosphorylates DAG to phosphatidic acid, which is the substrate used to produce 2-acyl lysophosphatidic acid (LPA) by the *LIPH* enzyme.

Importantly, signaling pathways that regulate lipid metabolism are involved in the regulation of hair growth, evidenced by mutations in the *P2RY5* and *LIPH* genes associated with the LAH and autosomal recessive woolly hair phenotypes in humans and mice (SONODA *et al.* 2002; PASTERNACK *et al.* 2008) (INOUE *et al.* 2011) (KAZANTSEVA *et al.* 2006; SHIMOMURA *et al.* 2008; SHIMOMURA *et al.* 2009). Although neither *Dkgz*$^{-/-}$ mice nor human patients with *DGKZ* or *DGKE* mutations displays an obvious hair phenotype (LEMAIRE *et al.* 2013) (OZALTIN *et al.* 2013), it is possible that a genetic interaction between the *DKGZ* and *ABCA5* genes produces the phenotype observed in autosomal recessive CGHT. Epistasis experiments using mouse models and functional studies in CGHT patient cells will clarify the contributions of each of these genes to disease pathogenesis.

Characterizing Polygenic Effects on Human Hair Growth

In the first two parts of my thesis, I examined the genetic basis of rare inherited conditions affecting hair growth, and using cytogenetic methods and next-generation sequencing, identified structural and intragenic variants underlying the sporadic and familial forms, respectively. In these disorders, the pathogenic variants are present at a very low frequency in the population (or *de novo*) and exert large effects, such that their causality to disease pathogenesis can often be clearly defined using the appropriate methods of detection and functional studies. However, the conditions affecting hair growth that are more common in the general population are typically composed of many genetic variants (polygenic) with small effects, whose contributions to disease pathogenesis are more difficult to characterize.

In Chapter IV of my thesis, I investigated the genetic architecture of a common, polygenic autoimmune disease of disfiguring hair loss, alopecia areata, and using a functional genomics approach, characterized the role of rare genetic variants at the *ULBP3/6* locus that is strongly implicated in disease pathogenesis. This locus encoding the NKG2D ligands was previously identified as the strongest association signal outside the *HLA* in the first GWAS for AA and demonstrated a strong upregulation of ULBP3 in human lesional hair follicles (PETUKHOVA *et al.* 2010). In Chapter IV.2, we performed a replication study and meta-GWAS analysis to expand gene discovery and identify true positive associations, which increased the genome-wide significance for the *ULBP3/6* locus association and also identified the BCL2L11 (BIM) gene as a novel locus associated with AA. Using several methods of detection, I characterized its expression in both human immune cells and hair follicles, and observed dysregulated Bim localization in mouse AA lesional hair follicles, implicating a role for this gene in disease pathogenesis. Based on the known functions of BIM as an apoptotic factor in immune cells and melanocytes, I postulate that BIM dysregulation in the AA hair follicle may contribute to the early entry into dystrophic catagen in AA hair follicles during active disease. Future follow-up studies will require functional analyses in mouse models, the identification of *BIM* susceptibility variants in human AA by targeted deep sequencing, and functional genomics to determine their consequences and contributions to disease pathogenesis.

In Chapter IV.3, I focused on the question, what are the genes that are dysregulated in alopecia areata? I characterized the transcriptional profile of human and mouse lesional skin and in collaboration with several lab members, uncovered three

primary gene expression signatures characteristic of immune cell activation and natural killer cell-mediated attack, further underscoring the NKG2D-ULBP signaling pathway. Furthermore, we demonstrated that T cells bearing the NKG2D receptor are both necessary and sufficient for disease pathogenesis (XING *et al.* 2014). Because these cells produce the cytokines and drivers of toxicity that are associated with the autoimmune response, we then targeted the downstream pathways of type I cytokines that signal through the JAK/STAT pathway using FDA-approved small molecule inhibitors (Ruxolitinib and Tofacitinib), and successfully prevented and reversed the disease in the mouse model. Subsequently, these findings led us to test the effectiveness of these small molecule inhibitors in human AA, and in a small pilot clinical trial, we successfully treated the disease and observed complete hair regrowth in the patients.

Collectively, our findings highlight the genetic and molecular basis of the collapse of immune privilege in the hair follicle that is characteristic of active disease in AA. Our genetic studies in human patients demonstrated strong associations with genes in both immune-related pathways (acting at the level of the effector T cells), and genes that are expressed in the target cells of the hair follicle. In lesional hair follicles, upregulated expression of NKG2D ligands in the dermal sheath cells is consistent with the observations that NKG2D+CD8T cells infiltrate the base of the hair follicle, and highlights the crosstalk between the target and effector cells in close proximity to one another, ultimately leading to cellular destruction. Importantly, these findings have defined the molecular basis of AA pathogenesis and shed light onto the previously uncharacterized cascade of signaling events during active disease in both humans and

mice. The requirement for NKG2D+CD8+ T cells for active disease, in conjunction with the genetic association of the NKG2D ligands with human AA, place the NKG2D-ULBP signaling axis squarely at the center of disease pathogenesis and invite the question of what genetic variants contribute to dysregulation of the *ULBP3/6* locus in human AA.

Functional Genomics Studies at the *ULBP6* Locus in AA

The quest to identify the pathogenic variants associated with complex human genetic disorders requires a robust and comprehensive approach integrating genetic, genomic, expression, and functional data. The functional genomics revolution was created to accomplish just this task, and designed to address the questions of what can we make of this information and how do we find the causal variants that disrupt regulatory and/or gene function. The genome-wide and high throughput integration of massive datasets allows us to tell a comprehensive biological story about the genetic, genomic, and functional information in a complex disease. Importantly, the ENCODE and Roadmap Epigenomics projects have greatly facilitated the process of interpreting mapped sequencing data through the systematic identification of functional regulatory elements including promoters, enhancers, and insulators. Notably, this approach, combined with CNV analysis, expression-based data and functional studies in mouse models is well powered to capture a large portion of the missing heritability in complex genetic disease.

In Chapter IV.4, the goal of my work was to identify and characterize the susceptibility variants at the *ULBP3/6* locus in AA that confer a degree of genetic risk in human AA. To accomplish this, we performed targeted deep resequencing of this locus

using the Raindance technology, and since the emerging picture of the genetic architecture underlying complex disease highlights a role of rare and non-coding variants, I focused my functional studies on a cluster of rare, novel variants at *ULBP6* that are enriched on the AA GWAS risk haplotypes. Using ENCODE and RegulomeDB, we found that these variants reside within regulatory regions, are predicted to disrupt CTCF sites downstream of *ULBP6*, and fall within paired-end tags from ENCODE ChIA-PET data on CTCF-mediated long-range chromatin interactions. Importantly, I demonstrated that CTCF is endogenously bound to the *ULBP6* candidate region in human dermal fibroblasts, and that the AA variants disrupt regulatory activity and CTCF binding *in vitro.* Using *in silico* analysis, I found that two of these variants create novel transcription factor binding sites for the activating proteins, ELK1 and EGR3, both of which have important roles in the immune system. Collectively, these findings support a role for rare, non-coding variants in disrupting the spatiotemporal regulation of the *ULBP3/6* locus in AA, and we postulate that these variants perturb long-range interactions that are otherwise required for *ULBP* repression in healthy tissues.

The findings from this study are exciting and novel, perfectly complementing the genetic studies in humans, immunopathological findings, and functional studies in the mouse model. In my analysis of the *ULBP3/6* locus, I exclusively tested the contribution of rare variants enriched on GWAS haplotypes under the assumption that they would produce larger effects than the common variants, however, I postulate that the combination of both rare and common variants on susceptibility haplotypes produce the maximal effects at this locus during active disease. In addition to CTCF, we identified other transcription factor binding site motifs in the *ULBP6* region with which the rare,

non-coding variants overlap (i.e. proteins of the cohesin and SUZ12 complexes), and I postulate that these factors also engage in long-range interactions that are potentially disrupted at this locus.

Until recently, the process of identifying and characterizing susceptibility variants in complex genetic diseases was very challenging. With thousands of SNPs to sift through (mostly in non-coding regions) for the hundreds of GWAS performed for common disorders, the primary limitation and bottleneck resided in the ability to identify the true positive associations and genomic regions harboring susceptibility variants that are actually pathogenic. Because the common-variant-common-disease assumption is intrinsic to the design of a GWAS, the success of uncovering variants that account for a significance portion of the heritability was limited, since we now know that rare variants and copy number variants (not detected by GWAS) significantly contribute to this "missing heritability" (GOLDSTEIN 2009; ZHU *et al.* 2011b). The identification of rare variants and their frequencies in complex genetic diseases will prove a challenging task, requiring the sequencing of many thousands of cases and controls, as rare variants are often not in LD with the common variants on risk haplotypes and may be population specific (ZUK *et al.* 2014). Since rare variants are likely to be deleterious and kept at a lower frequency by purifying selection, association tests to determine enrichment for individual rare variants may not be feasible, but rather, will require the aggregation of variants to be analyzed.

The large majority of the significant associations identified from GWAS studies reside within non-coding stretches of DNA several kilobases away from genes, highlighting that there is still much to be discovered in correlating genotype with

phenotype. However, several recent studies in complex traits have successfully implemented the functional genomics approach to characterize the function of non-coding variants and determine their impact on gene expression (i.e. expression quantitative trait loci (eQTL)). In the post-GWAS area, the correlation between sequence variation and mRNA expression was initially used to identify an eQTL, but more recent studies have examined variants in the context of TF binding, DNA methylation, chromatin accessibility, alternative splicing, lincRNAs, RNA editing, and mRNA degradation (ALBERT AND KRUGLYAK 2015).

Overlaying GWAS data for a given complex trait with eQTL mapping data in several tissues has proven a useful strategy to maximize the likelihood of finding GWAS-eQTL pairs, where this approach successfully detected a novel causal regulatory variant at the *NOS1AP* locus that modulates a heart-specific enhancer previously missed in eQTL studies performed on other tissues (KAPOOR *et al.* 2014). Similarly, GWAS variants in eQTLs have been shown to act in a long-range manner, as in the case of a common variant associated with myocardial infarction that resides within a 3' UTR but creates a novel TF binding site that upregulates the *SORT1* gene ~40 Kb away specifically in the liver. Remarkably, further functional studies in mouse demonstrated that SORT1 influences the levels of low-density lipoprotein cholesterol, a major risk factor for myocardial infarction (MUSUNURU *et al.* 2010). Moreover, in the context of hair-related complex genetic traits, a recent functional genomics study defined the molecular basis for classic blond hair in Europeans, elegantly demonstrating the mechanism by which a common SNP in the *KITLG* gene regulatory region influences hair follicle pigmentation in both humans and mice (GUENTHER *et al.* 2014).

Collectively, these findings demonstrate the applicability and success of the functional genomics approach to the study complex genetic diseases. In our human AA studies, we utilized a powerful and integrative functional genomics approach to decode the genetic variation at the *ULBP3/6* locus, and identified a cluster of rare, non-coding variants that disrupt the regulatory activity and transcription factor binding properties of this region, strongly supporting a role for these variants in disease pathogenesis.

Conclusion

In my thesis work, I have used three genetic conditions affecting hair growth as models to interrogate the molecular basis of the pathology and obtain novel insight into mechanisms and pathways required for normal hair growth. Importantly, the approaches I used to decode the genetic architecture of rare, sporadic and familial Mendelian, as well as common, polygenic conditions were instrumental in my success identifying complex structural variants, a splice site mutation in a novel gene associated with human hair growth, and rare, non-coding variants at the *ULBP3/6* locus in human disease. Thus, the work in this thesis has demonstrated that a combinatorial strategy employing the classical genetic approach, CNV analysis, targeted and genome-wide next-generation sequencing, as well as functional genomics studies proves the most successful for identifying disease-associated genes and novel genetic mechanisms that control our traits.

Perspectives

The completion of the human genome sequence greatly facilitated the development of

advanced technologies for whole-genome and targeted genome sequencing approaches, revolutionizing the field of genomics. In the context of health, reproductive fitness and personalized medicine, it is important to understand the underlying mechanisms and genes controlling our traits as well as the impacts of the surrounding environment. Over the past decade, we gained profound insight into the natural person-to-person variation and genetic heterogeneity at the DNA level that was previously unrecognized. Comparing the genomes of healthy individuals from distinct geographic regions shed light onto such sequence variation and the many changes that were tolerated, from SNPs to large insertions or deletions, raising an interesting question about what we consider pathogenic versus benign when searching for causal mutations in genetic diseases. Moreover, we learned that the coding portion of our DNA, or the exome, comprises a mere ~1% (30 Mb) of our genome, but its contribution to biological function is profound, evidenced by the large number of genetic diseases caused by mutations in single genes.

To date, mutations have been identified for approximately 50% of the 7,000 rare monogenic diseases reported in the OMIM database, and it is estimated that the remaining 50% will be identified by the year 2020. Recently, the Human Phenome Project has been launched to promote a collective integration of genetic and phenotypic data internationally that will facilitate the identification of novel disease genes and improve the definition and characterization of phenotypes. By interconnecting multiple databases and resources, cohorts can be combined to increase the statistical power for detection of novel variants and functionally decode the genotypes associated with the many human phenotypes.

In the context of common disease, deep sequencing technologies have afforded us a more comprehensive and detailed overview of the genetic variation within the regions of association identified by GWAS. Upon the targeted resequencing of GWAS cases, the focus in the field shifted from the common-variant-common-disease hypothesis to the current perspective, which supports a role for common variants with small effects, rare and structural variants with moderate effects, gene-gene interactions and gene-environment interactions. Although many studies and technological advances thereafter have greatly facilitated gene discovery in both rare and common disorders, there are still many unannotated genes and transcripts whose functions and contributions to disease pathogenesis remain unclear. Moreover, the massive amount of sequencing data generated from high-throughput experiments posed new challenges on how to process, analyze, and interpret this information. Over the past few years, there has been a surge in the development of novel bioinformatics tools and strategies aimed at accomplishing this task, and in the study of both rare and common diseases, a robust strategy integrating genotype and phenotype data with functional studies will undoubtedly provide deeper insights into the complexities of our genomic landscape, whose understanding is key for personalized genomics.

References

Adaimy, L., E. Chouery, H. Megarbane, S. Mroueh, V. Delague, E. Nicolas, H. Belguith, P. de Mazancourt and A. Megarbane, 2007 Mutation in WNT10A is associated with an autosomal recessive ectodermal dysplasia: the odonto-onycho-dermal dysplasia. Am J Hum Genet 81: 821-828.

Agarwal, S., J. P. Viola and A. Rao, 1999 Chromatin-based regulatory mechanisms governing cytokine gene transcription. J Allergy Clin Immunol 103: 990-999.

Ahmad, W., M. Faiyaz ul Haque, V. Brancolini, H. C. Tsou, S. ul Haque, H. Lam, V. M. Aita, J. Owen, M. deBlaquiere, J. Frank, P. B. Cserhalmi-Friedman, A. Leask, J. A. McGrath, M. Peacocke, M. Ahmad, J. Ott and A. M. Christiano, 1998 Alopecia universalis associated with a mutation in the human hairless gene. Science 279: 720-724.

Albert, F. W., and L. Kruglyak, 2015 The role of regulatory variation in complex traits and disease. Nat Rev Genet 16: 197-212.

Andl, T., S. T. Reddy, T. Gaddapara and S. E. Millar, 2002 WNT signals are required for the initiation of hair follicle development. Dev Cell 2: 643-653.

Androlewicz, M. J., K. S. Anderson and P. Cresswell, 1993 Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. Proc Natl Acad Sci U S A 90: 9130-9134.

Antonini, D., B. Rossi, R. Han, A. Minichiello, T. Di Palma, M. Corrado, S. Banfi, M. Zannini, J. L. Brissette and C. Missero, 2006 An autoregulatory loop directs the tissue-specific expression of p63 through a long-range evolutionarily conserved enhancer. Mol Cell Biol 26: 3308-3318.

Bahn, R. S., 2002 Thyrotropin receptor expression in orbital adipose/connective tissues from patients with thyroid-associated ophthalmopathy. Thyroid 12: 193-195.

Bahn, R. S., and A. E. Heufelder, 1993 Pathogenesis of Graves' ophthalmopathy. N Engl J Med 329: 1468-1475.

Balducci, R., V. Toscano, B. Tedeschi, A. Mangiantini, R. Toscano, C. Galasso, S. Cianfarani and B. Boscherini, 1998 A new case of Ambras syndrome associated with a paracentric inversion (8) (q12; q22). Clin Genet 53: 466-468.

Barahmani, N., M. de Andrade, J. P. Slusser, Q. Zhang and M. Duvic, 2006 Major histocompatibility complex class I chain-related gene A

polymorphisms and extended haplotypes are associated with familial alopecia areata. J Invest Dermatol 126: 74-78.

Barreto, O. C., 1998 [Hair transplantation. Past, present, and future]. Acta Med Port 11: 283-285.

Baumeister, F. A., J. Egger, M. T. Schildhauer and S. Stengel-Rutkowski, 1993 Ambras syndrome: delineation of a unique hypertrichosis universalis congenita and association with a balanced pericentric inversion (8) (p11.2; q22). Clin Genet 44: 121-128.

Beaudoin, M., P. Goyette, G. Boucher, K. S. Lo, M. A. Rivas, C. Stevens, A. Alikashani, M. Ladouceur, D. Ellinghaus, L. Torkvist, G. Goel, C. Lagace, V. Annese, A. Bitton, J. Begun, S. R. Brant, F. Bresso, J. H. Cho, R. H. Duerr, J. Halfvarson, D. P. McGovern, G. Radford-Smith, S. Schreiber, P. L. Schumm, Y. Sharma, M. S. Silverberg, R. K. Weersma, I. B. D. G. C. Quebec, N. I. G. Consortium, I. B. D. G. C. International, M. D'Amato, S. Vermeire, A. Franke, G. Lettre, R. J. Xavier, M. J. Daly and J. D. Rioux, 2013 Deep resequencing of GWAS loci identifies rare variants in CARD9, IL23R and RNF186 that are associated with ulcerative colitis. PLoS Genet 9: e1003723.

Beighton, P., 1970 Congenital hypertrichosis lanuginosa. Arch Dermatol 101: 669-672.

Bernheim, A., 2010 Cytogenomics of cancers: from chromosome to sequence. Mol Oncol 4: 309-322.

Betz, R. C., L. Petukhova, S. Ripke, H. Huang, A. Menelaou, S. Redler, T. Becker, S. Heilmann, T. Yamany, M. Duvic, M. Hordinsky, D. Norris, V. H. Price, J. Mackay-Wiggan, A. de Jong, G. M. DeStefano, S. Moebus, M. Bohm, U. Blume-Peytavi, H. Wolff, G. Lutz, R. Kruse, L. Bian, C. I. Amos, A. Lee, P. K. Gregersen, B. Blaumeiser, D. Altshuler, R. Clynes, P. I. de Bakker, M. M. Nothen, M. J. Daly and A. M. Christiano, 2015 Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. Nat Commun 6: 5966.

Biswas-Fiss, E. E., S. Affet, M. Ha and S. B. Biswas, 2012 Retinoid binding properties of nucleotide binding domain 1 of the Stargardt disease-associated ATP binding cassette (ABC) transporter, ABCA4. J Biol Chem 287: 44097-44107.

Borchers, M. T., S. C. Wesselkamper, V. Curull, A. Ramirez-Sarmiento, A. Sanchez-Font, J. Garcia-Aymerich, C. Coronell, J. Lloreta, A. G. Agusti, J. Gea, J. A. Howington, M. F. Reed, S. L. Starnes, N. L. Harris, M. Vitucci, B. L. Eppert, G. T. Motz, K. Fogel, D. W. McGraw, J. W. Tichelaar and M. Orozco-Levi, 2009 Sustained CTL activation by murine pulmonary epithelial cells promotes the development of COPD-like disease. J Clin Invest 119: 636-649.

Bowl, M. R., M. A. Nesbit, B. Harding, E. Levy, A. Jefferson, E. Volpi, K. Rizzoti, R. Lovell-Badge, D. Schlessinger, M. P. Whyte and R. V. Thakker, 2005 An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. J Clin Invest 115: 2822-2831.

Boycott, K. M., M. R. Vanstone, D. E. Bulman and A. E. MacKenzie, 2013 Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet 14: 681-691.

Cabral, R. M., M. Kurban, M. Wajid, Y. Shimomura, L. Petukhova and A. M. Christiano, 2012 Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. Genomics 99: 202-208.

Cadieu, E., M. W. Neff, P. Quignon, K. Walsh, K. Chase, H. G. Parker, B. M. Vonholdt, A. Rhue, A. Boyko, A. Byers, A. Wong, D. S. Mosher, A. G. Elkahloun, T. C. Spady, C. Andre, K. G. Lark, M. Cargill, C. D. Bustamante, R. K. Wayne and E. A. Ostrander, 2009 Coat variation in the domestic dog is governed by variants in three genes. Science 326: 150-153.

Caillat-Zucman, S., 2006 How NKG2D ligands trigger autoimmunity? Hum Immunol 67: 204-207.

Cantu, J. M., and C. Ruiz, 1985 On atavisms and atavistic genes. Ann Genet 28: 141-142.

Cantu, J. M., D. Garcia-Cruz, J. Sanchez-Corona, A. Hernandez and Z. Nazara, 1982a A distinct osteochondrodysplasia with hypertrichosis-Individualization of a probable autosomal recessive entity. Hum Genet 60: 36-41.

Cantu, J. M., J. Sanchez-Corona, A. Hernandes, Z. Nazara and D. Garcia-Cruz, 1982b Individualization of a syndrome with mental deficiency, macrocranium, peculiar facies, and cardiac and skeletal anomalies. Clin Genet 22: 172-179.

Cao, W., and W. He, 2004 UL16 binding proteins. Immunobiology 209: 283-290.

Celli, J., P. Duijf, B. C. Hamel, M. Bamshad, B. Kramer, A. P. Smits, R. Newbury-Ecob, R. C. Hennekam, G. Van Buggenhout, A. van Haeringen, C. G. Woods, A. J. van Essen, R. de Waal, G. Vriend, D. A. Haber, A. Yang, F. McKeon, H. G. Brunner and H. van Bokhoven, 1999 Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome. Cell 99: 143-153.

Chiao, E., P. Fisher, L. Crisponi, M. Deiana, I. Dragatsis, D. Schlessinger, G. Pilia and A. Efstratiadis, 2002 Overgrowth of a mouse model of the

Simpson-Golabi-Behmel syndrome is independent of IGF signaling. Dev Biol 243: 185-206.

Chien, A. J., M. C. Valentine and V. P. Sybert, 2006 Hereditary woolly hair and keratosis pilaris. J Am Acad Dermatol 54: S35-39.

Cirulli, E. T., and D. B. Goldstein, 2010 Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11: 415-425.

Clavel, C., L. Grisanti, R. Zemla, A. Rezza, R. Barros, R. Sennett, A. R. Mazloom, C. Y. Chung, X. Cai, C. L. Cai, L. Pevny, S. Nicolis, A. Ma'ayan and M. Rendl, 2012 Sox2 in the dermal papilla niche controls hair growth by fine-tuning BMP signaling in differentiating hair shaft progenitors. Dev Cell 23: 981-994.

Cluzeau, C., S. Hadj-Rabia, M. Jambou, S. Mansour, P. Guigue, S. Masmoudi, E. Bal, N. Chassaing, M. C. Vincent, G. Viot, F. Clauss, M. C. Maniere, S. Toupenay, M. Le Merrer, S. Lyonnet, V. Cormier-Daire, J. Amiel, L. Faivre, Y. de Prost, A. Munnich, J. P. Bonnefont, C. Bodemer and A. Smahi, 2011 Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. Hum Mutat 32: 70-72.

Contreras-Jurado, C., C. Lorz, L. Garcia-Serrano, J. M. Paramio and A. Aranda, 2015 Thyroid hormone signaling controls hair follicle stem cell function. Mol Biol Cell.

Cullinane, A. R., T. Vilboux, K. O'Brien, J. A. Curry, D. M. Maynard, H. Carlson-Donohoe, C. Ciccone, N. C. S. Program, T. C. Markello, M. Gunay-Aygun, M. Huizing and W. A. Gahl, 2011 Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia. J Invest Dermatol 131: 2017-2025.

Cunningham, M. D., J. A. Kassis and K. Pfeifer, 2010 Chromatin modifiers, cognitive disorders, and imprinted genes. Dev Cell 18: 169-170.

DeStefano, G. M., and A. M. Christiano, (In press) Basic Principles of Genetics. Genodermatoses, Section 9. Chapter 54 in Dermatology 4e.

DeStefano, G. M., and A. M. Christiano, 2014 The genetics of human skin disease. Cold Spring Harb Perspect Med 4.

DeStefano, G. M., K. A. Fantauzzo, L. Petukhova, M. Kurban, M. Tadin-Strapps, B. Levy, D. Warburton, E. T. Cirulli, Y. Han, X. Sun, Y. Shen, M. Shirazi, V. Jobanputra, R. Cepeda-Valdes, J. Cesar Salas-Alanis and A. M. Christiano, 2013 Position effect on FGF13 associated with X-linked congenital generalized hypertrichosis. Proc Natl Acad Sci U S A 110: 7790-7795.

DeStefano, G. M., M. Kurban, K. Anyane-Yeboa, C. Dall'Armi, G. Di Paolo, H. Feenstra, N. Silverberg, L. Rohena, L. D. Lopez-Cepeda, V. Jobanputra, K. A. Fantauzzo, M. Kiuru, M. Tadin-Strapps, A. Sobrino, A. Vitebsky, D. Warburton, B. Levy, J. C. Salas-Alanis and A. M. Christiano, 2014 Mutations in the cholesterol transporter gene ABCA5 are associated with excessive hair overgrowth. PLoS Genet 10: e1004333.

Di-Poi, N., L. Michalik, B. Desvergne and W. Wahli, 2004 Functions of peroxisome proliferator-activated receptors (PPAR) in skin homeostasis. Lipids 39: 1093-1099.

Dostie, J., T. A. Richmond, R. A. Arnaout, R. R. Selzer, W. L. Lee, T. A. Honan, E. D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R. D. Green and J. Dekker, 2006 Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16: 1299-1309.

Duboule, D., 1998 Vertebrate hox gene regulation: clustering and/or colinearity? Curr Opin Genet Dev 8: 514-518.

Dunn, S. M., R. A. Keough, G. E. Rogers and B. C. Powell, 1998 Regulation of a hair follicle keratin intermediate filament gene promoter. J Cell Sci 111 (Pt 23): 3487-3496.

Duverger, O., and M. I. Morasso, 2009 Epidermal patterning and induction of different hair types during mouse embryonic development. Birth Defects Res C Embryo Today 87: 263-272.

ENCODE, 2004 The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 306: 636-640.

Fantauzzo, K. A., and A. M. Christiano, 2012 Trps1 activates a network of secreted Wnt inhibitors and transcription factors crucial to vibrissa follicle morphogenesis. Development 139: 203-214.

Fantauzzo, K. A., M. Kurban, B. Levy and A. M. Christiano, 2012 Trps1 and its target gene Sox9 regulate epithelial proliferation in the developing hair follicle and are associated with hypertrichosis. PLoS Genet 8: e1003002.

Fantauzzo, K. A., M. Tadin-Strapps, Y. You, S. E. Mentzer, F. A. Baumeister, S. Cianfarani, L. Van Maldergem, D. Warburton, J. P. Sundberg and A. M. Christiano, 2008 A position effect on TRPS1 is associated with Ambras syndrome in humans and the Koala phenotype in mice. Hum Mol Genet 17: 3539-3551.

Farh, K. K., A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shoresh, H. Whitton, R. J. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos, P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler and B. E. Bernstein, 2015

Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518: 337-343.

Fernandez-Hojas, R., M. L. Huie, C. Navarro, C. Dominguez, M. Roig, D. Lopez-Coronas, S. Teijeira, K. Anyane-Yeboa and R. Hirschhorn, 2002 Identification of six novel mutations in the acid alpha-glucosidase gene in three Spanish patients with infantile onset glycogen storage disease type II (Pompe disease). Neuromuscul Disord 12: 159-166.

Fernandez, B. A., J. Siegel-Bartelt, J. A. Herbrick, I. Teshima and S. W. Scherer, 2005 Holoprosencephaly and cleidocranial dysplasia in a patient due to two position-effect mutations: case report and review of the literature. Clin Genet 68: 349-359.

Ferreira, S. I., E. Matoso, M. Pinto, J. Almeida, T. Liehr, J. B. Melo and I. M. Carreira, 2010 X-chromosome terminal deletion in a female with premature ovarian failure: Haploinsufficiency of X-linked genes as a possible explanation. Mol Cytogenet 3: 14.

Figuera, L. E., M. Pandolfo, P. W. Dunne, J. M. Cantu and P. I. Patel, 1995 Mapping of the congenital generalized hypertrichosis locus to chromosome Xq24-q27.1. Nat Genet 10: 202-207.

Frank, J., C. Pignata, A. A. Panteleyev, D. M. Prowse, H. Baden, L. Weiner, L. Gaetaniello, W. Ahmad, N. Pozzi, P. B. Cserhalmi-Friedman, V. M. Aita, H. Uyttendaele, D. Gordon, J. Ott, J. L. Brissette and A. M. Christiano, 1999 Exposing the human nude phenotype. Nature 398: 473-474.

Fryns, J. P., A. Kleczkowska and H. Kenis, 1984 De novo complex chromosomal rearrangement (CCR) in a severely mentally retarded boy. Ann Genet 27: 62-64.

Fu, Y., J. H. Hsiao, G. Paxinos, G. M. Halliday and W. S. Kim, 2015 ABCA5 regulates amyloid-beta peptide production and is associated with Alzheimer's disease neuropathology. J Alzheimers Dis 43: 857-869.

Fuchs, E., 2007 Scratching the surface of skin development. Nature 445: 834-842.

Fujimoto, A., M. Farooq, H. Fujikawa, A. Inoue, M. Ohyama, R. Ehama, J. Nakanishi, M. Hagihara, T. Iwabuchi, J. Aoki, M. Ito and Y. Shimomura, 2012 A missense mutation within the helix initiation motif of the keratin K71 gene underlies autosomal dominant woolly hair/hypotrichosis. J Invest Dermatol 132: 2342-2349.

Gandelman, M., and J. S. Epstein, 2004 Hair transplantation to the eyebrow, eyelashes, and other parts of the body. Facial Plast Surg Clin North Am 12: 253-261.

Gao, J., C. Liu, F. Yao, N. Hao, J. Zhou, Q. Zhou, L. Zhang, X. Liu, X. Bian and J. Liu, 2012 Array-based comparative genomic hybridization is more informative than conventional karyotyping and fluorescence in situ hybridization in the analysis of first-trimester spontaneous abortion. Mol Cytogenet 5: 33.

Garcia-Cruz, D., L. E. Figuera and J. M. Cantu, 2002 Inherited hypertrichoses. Clinical Genetics 61: 321-329.

Garn, S. M., 1951 Types and distribution of the hair in man. Ann N Y Acad Sci 53: 498-507.

Gat, U., R. DasGupta, L. Degenstein and E. Fuchs, 1998 De Novo hair follicle morphogenesis and hair tumors in mice expressing a truncated beta-catenin in skin. Cell 95: 605-614.

Gecz, J., E. Baker, A. Donnelly, J. E. Ming, D. M. McDonald-McGinn, N. B. Spinner, E. H. Zackai, G. R. Sutherland and J. C. Mulley, 1999 Fibroblast growth factor homologous factor 2 (FHF2): gene structure, expression and mapping to the Borjeson-Forssman-Lehmann syndrome region in Xq26 delineated by a duplication breakpoint in a BFLS-like patient. Hum Genet 104: 56-63.

Gharbi, S. I., E. Rincon, A. Avila-Flores, P. Torres-Ayuso, M. Almena, M. A. Cobos, J. P. Albar and I. Merida, 2011 Diacylglycerol kinase zeta controls diacylglycerol metabolism at the immunological synapse. Mol Biol Cell 22: 4406-4414.

Gilhar, A., A. Etzioni and R. Paus, 2012 Alopecia areata. N Engl J Med 366: 1515-1525.

Gilhar, A., and R. S. Kalish, 2006 Alopecia areata: a tissue specific autoimmune disease of the hair follicle. Autoimmun Rev 5: 64-69.

Gilhar, A., Y. Ullmann, T. Berkutzki, B. Assy and R. S. Kalish, 1998 Autoimmune hair loss (alopecia areata) transferred by T lymphocytes to human scalp explants on SCID mice. J Clin Invest 101: 62-67.

Gilissen, C., A. Hoischen, H. G. Brunner and J. A. Veltman, 2012 Disease gene identification strategies for exome sequencing. Eur J Hum Genet 20: 490-497.

Gillen, A. E., and A. Harris, 2011 The role of CTCF in coordinating the expression of single gene loci. Biochem Cell Biol 89: 489-494.

Goldstein, D. B., 2009 Common genetic variation and human traits. N Engl J Med 360: 1696-1698.

Goltz, R. W., 1992 Focal dermal hypoplasia syndrome. An update. Arch Dermatol 128: 1108-1111.

Gomez-Skarmeta, J. L., I. Rodriguez, C. Martinez, J. Culi, D. Ferres-Marco, D. Beamonte and J. Modolell, 1995 Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. Genes Dev 9: 1869-1882.

Gonzalez, S., V. Groh and T. Spies, 2006 Immunobiology of human NKG2D and its ligands. Curr Top Microbiol Immunol 298: 121-138.

Guenther, C. A., B. Tasic, L. Luo, M. A. Bedell and D. M. Kingsley, 2014 A molecular basis for classic blond hair color in Europeans. Nat Genet 46: 748-752.

Guibert, S., Z. Zhao, M. Sjolinder, A. Gondor, A. Fernandez, V. Pant and R. Ohlsson, 2012 CTCF-binding sites within the H19 ICR differentially regulate local chromatin structures and cis-acting functions. Epigenetics 7: 361-369.

Gusev, A., S. H. Lee, G. Trynka, H. Finucane, B. J. Vilhjalmsson, H. Xu, C. Zang, S. Ripke, B. Bulik-Sullivan, E. Stahl, C. Schizophrenia Working Group of the Psychiatric Genomics, S.-S. Consortium, A. K. Kahler, C. M. Hultman, S. M. Purcell, S. A. McCarroll, M. Daly, B. Pasaniuc, P. F. Sullivan, B. M. Neale, N. R. Wray, S. Raychaudhuri, A. L. Price, C. Schizophrenia Working Group of the Psychiatric Genomics and S.-S. Consortium, 2014 Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am J Hum Genet 95: 535-552.

Hall, B. K., 1984 Development mechanisms underlying the formation of atavisms. Biol Rev Camb Philos Soc 59: 89-124.

Hall, B. K., 1995 Atavisms and atavistic mutations. Nat Genet 10: 126-127.

Hall, B. K., 2010 Atavisms. Curr Biol 20: R871.

Harakalova, M., J. J. van Harssel, P. A. Terhal, S. van Lieshout, K. Duran, I. Renkens, D. J. Amor, L. C. Wilson, E. P. Kirk, C. L. Turner, D. Shears, S. Garcia-Minaur, M. M. Lees, A. Ross, H. Venselaar, G. Vriend, H. Takanari, M. B. Rook, M. A. van der Heyden, F. W. Asselbergs, H. M. Breur, M. E. Swinkels, I. J. Scurr, S. F. Smithson, N. V. Knoers, J. J. van der Smagt, I. J. Nijman, W. P. Kloosterman, M. M. van Haelst, G. van Haaften and E. Cuppen, 2012 Dominant missense mutations in ABCC9 cause Cantu syndrome. Nat Genet 44: 793-796.

Hebert, J. M., T. Rosenquist, J. Gotz and G. R. Martin, 1994 FGF5 as a regulator of the hair growth cycle: evidence from targeted and spontaneous mutations. Cell 78: 1017-1025.

Heimerl, S., A. K. Bosserhoff, T. Langmann, J. Ecker and G. Schmitz, 2007 Mapping ATP-binding cassette transporter gene expression profiles in melanocytes and melanoma cells. Melanoma Res 17: 265-273.

Held, L. I., Jr., 2010 The Evo-Devo Puzzle of Human Hair Patterning. Evol Biol 37: 113–122.

Henderson, R. A., A. L. Cox, K. Sakaguchi, E. Appella, J. Shabanowitz, D. F. Hunt and V. H. Engelhard, 1993 Direct identification of an endogenous peptide recognized by multiple HLA-A2.1-specific cytotoxic T cells. Proc Natl Acad Sci U S A 90: 10275-10279.

Hesse, M., A. Zimek, K. Weber and T. M. Magin, 2004 Comprehensive analysis of keratin gene clusters in humans and rodents. Eur J Cell Biol 83: 19-26.

Higgins, C. A., L. Petukhova, S. Harel, Y. Y. Ho, E. Drill, L. Shapiro, M. Wajid and A. M. Christiano, 2014 FGF5 is a crucial regulator of hair length in humans. Proc Natl Acad Sci U S A 111: 10648-10653.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins and T. A. Manolio, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362-9367.

Hnisz, D., B. J. Abraham, T. I. Lee, A. Lau, V. Saint-Andre, A. A. Sigova, H. A. Hoke and R. A. Young, 2013 Super-enhancers in the control of cell identity and disease. Cell 155: 934-947.

Hoischen, A., B. W. van Bon, C. Gilissen, P. Arts, B. van Lier, M. Steehouwer, P. de Vries, R. de Reuver, N. Wieskamp, G. Mortier, K. Devriendt, M. Z. Amorim, N. Revencu, A. Kidd, M. Barbosa, A. Turner, J. Smith, C. Oley, A. Henderson, I. M. Hayes, E. M. Thompson, H. G. Brunner, B. B. de Vries and J. A. Veltman, 2010 De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet 42: 483-485.

Horike, S., S. Cai, M. Miyano, J. F. Cheng and T. Kohwi-Shigematsu, 2005 Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet 37: 31-40.

Hutchinson, J. N., T. Raj, J. Fagerness, E. Stahl, F. T. Viloria, A. Gimelbrant, J. Seddon, M. Daly, A. Chess and R. Plenge, 2014 Allele-specific methylation occurs at genetic variants associated with complex disease. PLoS One 9: e98464.

Imcke, E., A. Mayer-da-Silva, M. Detmar, H. Tiel, R. Stadler and C. E. Orfanos, 1987 Growth of human hair follicle keratinocytes in vitro. Ultrastructural features of a new model. J Am Acad Dermatol 17: 779-786.

Inoue, A., N. Arima, J. Ishiguro, G. D. Prestwich, H. Arai and J. Aoki, 2011 LPA-producing enzyme PA-PLA(1)alpha regulates hair follicle development by modulating EGFR signalling. EMBO J 30: 4248-4260.

Ito, T., K. C. Meyer, N. Ito and R. Paus, 2008b Immune privilege and the skin. Curr Dir Autoimmun 10: 27-52.

Ito, T., N. Ito, M. Saatoff, H. Hashizume, H. Fukamizu, B. J. Nickoloff, M. Takigawa and R. Paus, 2008 Maintenance of hair follicle immune privilege is linked to prevention of NK cell attack. J Invest Dermatol 128: 1196-1206.

Ito, T., N. Ito, M. Saatoff, H. Hashizume, H. Fukamizu, B. J. Nickoloff, M. Takigawa and R. Paus, 2008a Maintenance of hair follicle immune privilege is linked to prevention of NK cell attack. J Invest Dermatol 128: 1196-1206.

Jackow, C., N. Puffer, M. Hordinsky, J. Nelson, J. Tarrand and M. Duvic, 1998 Alopecia areata and cytomegalovirus infection in twins: genes versus environment? J Am Acad Dermatol 38: 418-425.

Jakubovic, B. D., and S. Jothy, 2007 Glypican-3: from the mutations of Simpson-Golabi-Behmel genetic syndrome to a tumor marker for hepatocellular carcinoma. Exp Mol Pathol 82: 184-189.

Jhunjhunwala, S., M. C. van Zelm, M. M. Peak and C. Murre, 2009 Chromatin architecture and the generation of antigen receptor diversity. Cell 138: 435-448.

Kadaja, M., B. E. Keyes, M. Lin, H. A. Pasolli, M. Genander, L. Polak, N. Stokes, D. Zheng and E. Fuchs, 2014 SOX9: a stem cell transcriptional regulator of secreted niche signaling factors. Genes Dev 28: 328-341.

Kang, S. H., C. Shaw, Z. Ou, P. A. Eng, M. L. Cooper, A. N. Pursley, T. Sahoo, C. A. Bacino, A. C. Chinault, P. Stankiewicz, A. Patel, J. R. Lupski and S. W. Cheung, 2010 Insertional translocation detected using FISH confirmation of array-comparative genomic hybridization (aCGH) results. Am J Med Genet A 152A: 1111-1126.

Kapoor, A., R. B. Sekar, N. F. Hansen, K. Fox-Talbot, M. Morley, V. Pihur, S. Chatterjee, J. Brandimarto, C. S. Moravec, S. L. Pulit, Q. T. I.-I. G. Consortium, A. Pfeufer, J. Mullikin, M. Ross, E. D. Green, D. Bentley, C. Newton-Cheh, E. Boerwinkle, G. F. Tomaselli, T. P. Cappola, D. E. Arking, M. K. Halushka and A. Chakravarti, 2014 An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. Am J Hum Genet 94: 854-869.

Karnik, P., Z. Tekeste, T. S. McCormick, A. C. Gilliam, V. H. Price, K. D. Cooper and P. Mirmirani, 2009 Hair follicle stem cell-specific PPARgamma deletion causes scarring alopecia. J Invest Dermatol 129: 1243-1257.

Kawano, M., S. Suzuki, M. Suzuki, J. Oki and T. Imamura, 2004 Bulge- and basal layer-specific expression of fibroblast growth factor-13 (FHF-2) in mouse skin. J Invest Dermatol 122: 1084-1090.

Kazantseva, A., A. Goltsov, R. Zinchenko, A. P. Grigorenko, A. V. Abrukova, Y. K. Moliaka, A. G. Kirillov, Z. Guo, S. Lyle, E. K. Ginter and E. I. Rogaev, 2006 Human hair growth deficiency is linked to a genetic defect in the phospholipase gene LIPH. Science 314: 982-985.

Kleinjan, D. J., and V. van Heyningen, 1998 Position effect in human genetic disease. Hum Mol Genet 7: 1611-1618.

Kljuic, A., H. Bazzi, J. P. Sundberg, A. Martinez-Mir, R. O'Shaughnessy, M. G. Mahoney, M. Levy, X. Montagutelli, W. Ahmad, V. M. Aita, D. Gordon, J. Uitto, D. Whiting, J. Ott, S. Fischer, T. C. Gilliam, C. A. Jahoda, R. J. Morris, A. A. Panteleyev, V. T. Nguyen and A. M. Christiano, 2003 Desmoglein 4 in hair follicle differentiation and epidermal adhesion: evidence from inherited hypotrichosis and acquired pemphigus vulgaris. Cell 113: 249-260.

Ko, T. M., W. L. Hwu, Y. W. Lin, L. H. Tseng, H. L. Hwa, T. R. Wang and S. M. Chuang, 1999 Molecular genetic study of Pompe disease in Chinese patients in Taiwan. Hum Mutat 13: 380-384.

Kubo, Y., S. Sekiya, M. Ohigashi, C. Takenaka, K. Tamura, S. Nada, T. Nishi, A. Yamamoto and A. Yamaguchi, 2005 ABCA5 resides in lysosomes, and ABCA5 knockout mice develop lysosomal disease-like symptoms. Mol Cell Biol 25: 4138-4149.

Kurban, M., C. A. Kim, M. Kiuru, K. Fantauzzo, R. Cabral, O. Abbas, B. Levy and A. M. Christiano, 2011 Copy number variations on chromosome 4q26-27 are associated with Cantu syndrome. Dermatology 223: 316-320.

Lai-Cheong, J. E., and J. A. McGrath, 2010 Next-generation diagnostics for inherited skin disorders. J Invest Dermatol 131: 1971-1973.

LeBlanc, S. E., Q. Wu, A. R. Barutcu, H. Xiao, Y. Ohkawa and A. N. Imbalzano, 2014 The PPARgamma locus makes long-range chromatin interactions with selected tissue-specific gene loci during adipocyte differentiation in a protein kinase A dependent manner. PLoS One 9: e86140.

Lee, G. R., C. G. Spilianakis and R. A. Flavell, 2005 Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. Nat Immunol 6: 42-48.

Lee, S. H., D. W. Kim, J. B. Jun, S. J. Lee, J. C. Kim and N. H. Kim, 1999 The changes in hair growth pattern after autologous hair transplantation. Dermatol Surg 25: 605-609.

Leipoldt, M., M. Erdel, G. A. Bien-Willner, M. Smyk, M. Theurl, S. A. Yatsenko, J. R. Lupski, A. H. Lane, A. L. Shanske, P. Stankiewicz and G. Scherer, 2007 Two novel translocation breakpoints upstream of SOX9 define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. Clin Genet 71: 67-75.

Lemaire, M., V. Fremeaux-Bacchi, F. Schaefer, M. Choi, W. H. Tang, M. Le Quintrec, F. Fakhouri, S. Taque, F. Nobili, F. Martinez, W. Ji, J. D. Overton, S. M. Mane, G. Nurnberg, J. Altmuller, H. Thiele, D. Morin, G. Deschenes, V. Baudouin, B. Llanas, L. Collard, M. A. Majid, E. Simkova, P. Nurnberg, N. Rioux-Leclerc, G. W. Moeckel, M. C. Gubler, J. Hwa, C. Loirat and R. P. Lifton, 2013 Recessive mutations in DGKE cause atypical hemolytic-uremic syndrome. Nat Genet 45: 531-536.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill and E. de Graaff, 2003 A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12: 1725-1735.

Li, G., M. J. Fullwood, H. Xu, F. H. Mulawadi, S. Velkov, V. Vega, P. N. Ariyaratne, Y. B. Mohamed, H. S. Ooi, C. Tennakoon, C. L. Wei, Y. Ruan and W. K. Sung, 2010 ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol 11: R22.

Li, M., C. Shuman, Y. L. Fei, E. Cutiongco, H. A. Bender, C. Stevens, L. Wilkins-Haug, D. Day-Salvatore, S. L. Yong, M. T. Geraghty, J. Squire and R. Weksberg, 2001 GPC3 mutation analysis in a spectrum of patients with overgrowth expands the phenotype of Simpson-Golabi-Behmel syndrome. Am J Med Genet 102: 161-168.

Li, S., T. Miao, M. Sebastian, P. Bhullar, E. Ghaffari, M. Liu, A. L. Symonds and P. Wang, 2012 The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells. Immunity 37: 685-696.

Li, T., J. F. Hu, X. Qiu, J. Ling, H. Chen, S. Wang, A. Hou, T. H. Vu and A. R. Hoffman, 2008 CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. Mol Cell Biol 28: 6473-6482.

Lindhurst, M. J., J. C. Sapp, J. K. Teer, J. J. Johnston, E. M. Finn, K. Peters, J. Turner, J. L. Cannons, D. Bick, L. Blakemore, C. Blumhorst, K. Brockmann, P. Calder, N. Cherman, M. A. Deardorff, D. B. Everman, G. Golas, R. M. Greenstein, B. M. Kato, K. M. Keppler-Noreuil, S. A. Kuznetsov, R. T. Miyamoto, K. Newman, D. Ng, K. O'Brien, S. Rothenberg, D. J. Schwartzentruber, V. Singhal, R. Tirabosco, J. Upton, S. Wientroub, E. H. Zackai, K. Hoag, T. Whitewood-Neal, P. G. Robey, P. L. Schwartzberg, T. N. Darling, L. L. Tosi, J. C. Mullikin and L. G. Biesecker, 2011 A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med 365: 611-619.

Liu, X., J. Wang and L. Chen, 2013 Whole-exome sequencing reveals recurrent somatic mutation networks in cancer. Cancer Lett 340: 270-276.

Lohr, J. G., P. Stojanov, M. S. Lawrence, D. Auclair, B. Chapuy, C. Sougnez, P. Cruz-Gordillo, B. Knoechel, Y. W. Asmann, S. L. Slager, A. J. Novak, A. Dogan, S. M. Ansell, B. K. Link, L. Zou, J. Gould, G. Saksena, N. Stransky, C. Rangel-Escareno, J. C. Fernandez-Lopez, A. Hidalgo-Miranda, J. Melendez-Zajgla, E. Hernandez-Lemus, A. Schwarz-Cruz y Celis, I. Imaz-Rosshandler, A. I. Ojesina, J. Jung, C. S. Pedamallu, E. S. Lander, T. M. Habermann, J. R. Cerhan, M. A. Shipp, G. Getz and T. R. Golub, 2012 Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A 109: 3879-3884.

Luo, B., S. M. Prescott and M. K. Topham, 2004 Diacylglycerol kinase zeta regulates phosphatidylinositol 4-phosphate 5-kinase Ialpha by a novel mechanism. Cell Signal 16: 891-897.

Luo, S., and D. C. Rubinsztein, 2013 BCL2L11/BIM: a novel molecular link between autophagy and apoptosis. Autophagy 9: 104-105.

Lupski, J. R., J. G. Reid, C. Gonzaga-Jauregui, D. Rio Deiros, D. C. Chen, L. Nazareth, M. Bainbridge, H. Dinh, C. Jing, D. A. Wheeler, A. L. McGuire, F. Zhang, P. Stankiewicz, J. J. Halperin, C. Yang, C. Gehman, D. Guo, R. K. Irikat, W. Tom, N. J. Fantin, D. M. Muzny and R. A. Gibbs, 2010 Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 362: 1181-1191.

MacArthur, D. G., T. A. Manolio, D. P. Dimmock, H. L. Rehm, J. Shendure, G. R. Abecasis, D. R. Adams, R. B. Altman, S. E. Antonarakis, E. A. Ashley, J. C. Barrett, L. G. Biesecker, D. F. Conrad, G. M. Cooper, N. J. Cox, M. J. Daly, M. B. Gerstein, D. B. Goldstein, J. N. Hirschhorn, S. M. Leal, L. A. Pennacchio, J. A. Stamatoyannopoulos, S. R. Sunyaev, D. Valle, B. F. Voight, W. Winckler and C. Gunter, 2014 Guidelines for investigating causality of sequence variants in human disease. Nature 508: 469-476.

Majumder, P., and J. M. Boss, 2010 CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. Mol Cell Biol 30: 4211-4223.

Majumder, P., J. A. Gomez, B. P. Chadwick and J. M. Boss, 2008 The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. J Exp Med 205: 785-798.

Mallo, M., D. M. Wellik and J. Deschamps, 2010 Hox genes and regional patterning of the vertebrate body plan. Dev Biol 344: 7-15.

Manganelli, F., and L. Ruggiero, 2013 Clinical features of Pompe disease. Acta Myol 32: 82-84.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H.

Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll and P. M. Visscher, 2009 Finding the missing heritability of complex diseases. Nature 461: 747-753.

Mao, L. M., Q. Tang and J. Q. Wang, 2007 Protein kinase C-regulated cAMP response element-binding protein phosphorylation in cultured rat striatal neurons. Brain Res Bull 72: 302-308.

Mardaryev, A. N., M. R. Gdula, J. L. Yarker, V. U. Emelianov, K. Poterlowicz, A. A. Sharov, T. Y. Sharova, J. A. Scarpa, B. Joffe, I. Solovei, P. Chambon, V. A. Botchkarev and M. Y. Fessing, 2014 p63 and Brg1 control developmentally regulated higher-order chromatin remodelling at the epidermal differentiation complex locus in epidermal progenitor cells. Development 141: 101-111.

Mardis, E. R., 2008 The impact of next-generation sequencing technology on genetics. Trends Genet 24: 133-141.

Markiewicz, M. A., E. L. Wise, Z. S. Buchwald, A. K. Pinto, B. Zafirova, B. Polic and A. S. Shaw, 2012 RAE1epsilon ligand expressed on pancreatic islets recruits NKG2D receptor-expressing cytotoxic T cells independent of T cell receptor recognition. Immunity 36: 132-141.

Martinez-Mir, A., A. Zlotogorski, D. Gordon, L. Petukhova, J. Mo, T. C. Gilliam, D. Londono, C. Haynes, J. Ott, M. Hordinsky, K. Nanova, D. Norris, V. Price, M. Duvic and A. M. Christiano, 2007 Genomewide scan for linkage reveals evidence of several susceptibility loci for alopecia areata. Am J Hum Genet 80: 316-328.

Martinez-Mir, A., A. Zlotogorski, J. Ott, D. Gordon and A. M. Christiano, 2003 Genetic linkage studies in alopecia areata. J Investig Dermatol Symp Proc 8: 199-203.

Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul and J. A. Stamatoyannopoulos, 2012 Systematic localization of common disease-associated variation in regulatory DNA. Science 337: 1190-1195.

McDonagh, A. J., and R. Tazi-Ahnini, 2002 Epidemiology and genetics of alopecia areata. Clin Exp Dermatol 27: 405-409.

McElwee, K. J., D. Boggess, L. E. King, Jr. and J. P. Sundberg, 1998 Experimental induction of alopecia areata-like hair loss in C3H/HeJ mice using full-thickness skin grafts. J Invest Dermatol 111: 797-803.

Mehta, K. D., A. Radominska-Pandya, G. S. Kapoor, B. Dave and B. A. Atkins, 2002 Critical role of diacylglycerol- and phospholipid-regulated protein kinase C epsilon in induction of low-density lipoprotein receptor transcription in response to depletion of cholesterol. Mol Cell Biol 22: 3783-3793.

Millar, S. E., 2002 Molecular mechanisms regulating hair follicle development. Journal of Investigative Dermatology 118: 216-225.

Miller, O. J., Therman, E. , 2001 Human Chromosomes. Springer-Verlag, New York.

Moore, G., 1965 Cramming more components onto integrated circuits. Electronics 38.

Morozova, O., and M. A. Marra, 2008 Applications of next-generation sequencing technologies in functional genomics. Genomics 92: 255-264.

Morton, N. E., 1955 Sequential tests for the detection of linkage. Am J Hum Genet 7: 277-318.

Musio, A., A. Selicorni, M. L. Focarelli, C. Gervasini, D. Milani, S. Russo, P. Vezzoni and L. Larizza, 2006 X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet 38: 528-530.

Musunuru, K., A. Strong, M. Frank-Kamenetsky, N. E. Lee, T. Ahfeldt, K. V. Sachs, X. Li, H. Li, N. Kuperwasser, V. M. Ruda, J. P. Pirruccello, B. Muchmore, L. Prokunina-Olsson, J. L. Hall, E. E. Schadt, C. R. Morales, S. Lund-Katz, M. C. Phillips, J. Wong, W. Cantley, T. Racie, K. G. Ejebe, M. Orho-Melander, O. Melander, V. Koteliansky, K. Fitzgerald, R. M. Krauss, C. A. Cowan, S. Kathiresan and D. J. Rader, 2010 From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature 466: 714-719.

Nag, D. K., and A. Kurst, 1997 A 140-bp-long palindromic sequence induces double-strand breaks during meiosis in the yeast Saccharomyces cerevisiae. Genetics 146: 835-847.

Nakamura, M., M. R. Schneider, R. Schmidt-Ullrich and R. Paus, 2013 Mutant laboratory mice with abnormalities in hair follicle morphogenesis, cycling, and/or structure: an update. J Dermatol Sci 69: 6-29.

Nativio, R., A. Sparago, Y. Ito, R. Weksberg, A. Riccio and A. Murrell, 2011 Disruption of genomic neighbourhood at the imprinted IGF2-H19 locus in Beckwith-Wiedemann syndrome and Silver-Russell syndrome. Hum Mol Genet 20: 1363-1374.

Nativio, R., K. S. Wendt, Y. Ito, J. E. Huddleston, S. Uribe-Lewis, K. Woodfine, C. Krueger, W. Reik, J. M. Peters and A. Murrell, 2009 Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet 5: e1000739.

Nowak, J. A., L. Polak, H. A. Pasolli and E. Fuchs, 2008 Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell 3: 33-43.

O'Reilly, L. A., L. Cullen, J. Visvader, G. J. Lindeman, C. Print, M. L. Bath, D. C. Huang and A. Strasser, 2000 The proapoptotic BH3-only protein bim is expressed in hematopoietic, epithelial, neuronal, and germ cells. Am J Pathol 157: 449-461.

Oh, I. Y., D. M. Albea, Z. A. Goodwin, A. M. Quiggle, B. P. Baker, A. M. Guggisberg, J. H. Geahlen, G. M. Kroner and C. de Guzman Strong, 2014 Regulation of the dynamic chromatin architecture of the epidermal differentiation complex is mediated by a c-Jun/AP-1-modulated enhancer. J Invest Dermatol 134: 2371-2380.

Ohtsuki, S., M. Kamoi, Y. Watanabe, H. Suzuki, S. Hori and T. Terasaki, 2007 Correlation of induction of ATP binding cassette transporter A5 (ABCA5) and ABCB1 mRNAs with differentiation state of human colon tumor. Biol Pharm Bull 30: 1144-1146.

Ohyama, M., 2007 Hair follicle bulge: a fascinating reservoir of epithelial stem cells. J Dermatol Sci 46: 81-89.

Orentreich, N., 1959 Autografts in alopecias and other selected dermatological conditions. Ann N Y Acad Sci 83: 463-479.

Orr, M. B., R. M. Barlow and M. L. Ryder, 1977 Histological and histochemical studies of fetal skin follicles in Border disease of sheep. Res Vet Sci 22: 56-61.

Osborne, C. S., L. Chakalova, K. E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J. A. Mitchell, S. Lopes, W. Reik and P. Fraser, 2004 Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36: 1065-1071.

Ozaltin, F., B. Li, A. Rauhauser, S. W. An, O. Soylemezoglu, Gonul, II, E. Z. Taskiran, T. Ibsirlioglu, E. Korkmaz, Y. Bilginer, A. Duzova, S. Ozen, R. Topaloglu, N. Besbas, S. Ashraf, Y. Du, C. Liang, P. Chen, D. Lu, K. Vadnagara, S. Arbuckle, D. Lewis, B. Wakeland, R. J. Quigg, R. F. Ransom, E. K. Wakeland, M. K. Topham, N. G. Bazan, C. Mohan, F. Hildebrandt, A. Bakkaloglu, C. L. Huang and M. Attanasio, 2013 DGKE variants cause a glomerular microangiopathy that mimics membranoproliferative GN. J Am Soc Nephrol 24: 377-384.

Pabinger, S., A. Dander, M. Fischer, R. Snajder, M. Sperk, M. Efremova, B. Krabichler, M. R. Speicher, J. Zschocke and Z. Trajanoski, 2014 A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform 15: 256-278.

Pacheco, M. A., and R. S. Jope, 1996 Phosphoinositide signaling in human brain. Prog Neurobiol 50: 255-273.

Panteleyev, A. A., N. V. Botchkareva, J. P. Sundberg, A. M. Christiano and R. Paus, 1999 The role of the hairless (hr) gene in the regulation of hair follicle catagen transformation. Am J Pathol 155: 159-171.

Pasternack, S. M., I. von Kugelgen, K. Al Aboud, Y. A. Lee, F. Ruschendorf, K. Voss, A. M. Hillmer, G. J. Molderings, T. Franz, A. Ramirez, P. Nurnberg, M. M. Nothen and R. C. Betz, 2008 G protein-coupled receptor P2Y5 and its ligand LPA are involved in maintenance of human hair growth. Nat Genet 40: 329-334.

Petry, F., A. Kotthaus and K. I. Hirsch-Ernst, 2003 Cloning of human and rat ABCA5/Abca5 and detection of a human splice variant. Biochem Biophys Res Commun 300: 343-350.

Petry, F., V. Ritz, C. Meineke, P. Middel, T. Kietzmann, C. Schmitz-Salue and K. I. Hirsch-Ernst, 2006 Subcellular localization of rat Abca5, a rat ATP-binding-cassette transporter expressed in Leydig cells, and characterization of its splice variant apparently encoding a half-transporter. Biochem J 393: 79-87.

Petukhova, L., and A. M. Christiano, 2013 The genetic architecture of alopecia areata. J Investig Dermatol Symp Proc 16: S16-22.

Petukhova, L., Duvic, M., Hordinsky, M., Norris, D., Price, V., Shimomura, Y., Kim, H., Singh, P., Lee, A., Chen, W.V., Meyer, K.C., Paus, R., Jahoda, C.A.B, Amos, C.I., Gregersen, P.K., Christiano, A.M., 2010 Genome-wide Association Study in Alopecia Areata Implicates both Innate and Adaptive Immunity, pp. in Society for Investigative Dermatology 70th Annual Meeting, Atlanta, GA.

Petukhova, L., M. Duvic, M. Hordinsky, D. Norris, V. Price, Y. Shimomura, H. Kim, P. Singh, A. Lee, W. V. Chen, K. C. Meyer, R. Paus, C. A. Jahoda, C. I. Amos, P. K. Gregersen and A. M. Christiano, 2010 Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature 466: 113-117.

Petukhova, L., Y. Shimomura, M. Wajid, P. Gorroochurn, S. E. Hodge and A. M. Christiano, 2009 The effect of inbreeding on the distribution of compound heterozygotes: a lesson from Lipase H mutations in autosomal recessive woolly hair/hypotrichosis. Hum Hered 68: 117-130.

Phillips, J. E., and V. G. Corces, 2009 CTCF: master weaver of the genome. Cell 137: 1194-1211.

Pie, J., M. C. Gil-Rodriguez, M. Ciero, E. Lopez-Vinas, M. P. Ribate, M. Arnedo, M. A. Deardorff, B. Puisac, J. Legarreta, J. C. de Karam, E. Rubio, I. Bueno, A. Baldellou, M. T. Calvo, N. Casals, J. L. Olivares, A. Losada, F. G. Hegardt, I. D. Krantz, P. Gomez-Puertas and F. J. Ramos, 2010 Mutations and variants in the cohesion factor genes NIPBL, SMC1A, and SMC3 in a cohort of 30 unrelated patients with Cornelia de Lange syndrome. Am J Med Genet A 152A: 924-929.

Pilia, G., R. M. Hughes-Benzie, A. MacKenzie, P. Baybayan, E. Y. Chen, R. Huber, G. Neri, A. Cao, A. Forabosco and D. Schlessinger, 1996 Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. Nat Genet 12: 241-247.

Pinkel, D., and D. G. Albertson, 2005 Comparative genomic hybridization. Annu Rev Genomics Hum Genet 6: 331-354.

Powell, B. C., A. Nesci and G. E. Rogers, 1991 Regulation of keratin gene expression in hair follicle differentiation. Ann N Y Acad Sci 642: 1-20.

Powell, B. C., and G. E. Rogers, 1997 The role of keratin proteins and their genes in the growth, structure and properties of hair. EXS 78: 59-148.

Pugh, T. J., S. D. Weeraratne, T. C. Archer, D. A. Pomeranz Krummel, D. Auclair, J. Bochicchio, M. O. Carneiro, S. L. Carter, K. Cibulskis, R. L. Erlich, H. Greulich, M. S. Lawrence, N. J. Lennon, A. McKenna, J. Meldrim, A. H. Ramos, M. G. Ross, C. Russ, E. Shefler, A. Sivachenko, B. Sogoloff, P. Stojanov, P. Tamayo, J. P. Mesirov, V. Amani, N. Teider, S. Sengupta, J. P. Francois, P. A. Northcott, M. D. Taylor, F. Yu, G. R. Crabtree, A. G. Kautzman, S. B. Gabriel, G. Getz, N. Jager, D. T. Jones, P. Lichter, S. M. Pfister, T. M. Roberts, M. Meyerson, S. L. Pomeroy and Y. J. Cho, 2012 Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 488: 106-110.

Qazi, Q. H., L. S. Heckman, D. Markouizos and R. S. Verma, 1990 The Coffin-Siris syndrome. J Med Genet 27: 333-336.

Quazi, F., and R. S. Molday, 2011 Lipid transport by mammalian ABC proteins. Essays Biochem 50: 265-290.

Tadin-Strapps, M., J. C. Salas-Alanis, L. Moreno, D. Warburton, A. Martinez-Mir and A. M. Christiano, 2003 Congenital universal hypertrichosis with deafness and dental anomalies inherited as an X-linked trait. Clin Genet 63: 418-422.

Quesada, V., L. Conde, N. Villamor, G. R. Ordonez, P. Jares, L. Bassaganyas, A. J. Ramsay, S. Bea, M. Pinyol, A. Martinez-Trillos, M. Lopez-Guerra, D.

Colomer, A. Navarro, T. Baumann, M. Aymerich, M. Rozman, J. Delgado, E. Gine, J. M. Hernandez, M. Gonzalez-Diaz, D. A. Puente, G. Velasco, J. M. Freije, J. M. Tubio, R. Royo, J. L. Gelpi, M. Orozco, D. G. Pisano, J. Zamora, M. Vazquez, A. Valencia, H. Himmelbauer, M. Bayes, S. Heath, M. Gut, I. Gut, X. Estivill, A. Lopez-Guillermo, X. S. Puente, E. Campo and C. Lopez-Otin, 2012 Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44: 47-52.

Vidal, V. P., M. C. Chaboissier, S. Lutzkendorf, G. Cotsarelis, P. Mill, C. C. Hui, N. Ortonne, J. P. Ortonne and A. Schedl, 2005 Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Curr Biol 15: 1340-1351.

Ramot, Y., A. Mastrofrancesco, E. Camera, P. Desreumaux, R. Paus and M. Picardo, 2015 The role of PPARgamma-mediated signalling in skin biology and pathology: New targets and opportunities for clinical dermatology. Exp Dermatol.

Richardson, P., W. McKenna, M. Bristow, B. Maisch, B. Mautner, J. O'Connell, E. Olsen, G. Thiene, J. Goodwin, I. Gyarfas, I. Martin and P. Nordet, 1996 Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of cardiomyopathies. Circulation 93: 841-842.

Rincon, E., S. I. Gharbi, T. Santos-Mendoza and I. Merida, 2012 Diacylglycerol kinase zeta: at the crossroads of lipid signaling and protein complex organization. Prog Lipid Res 51: 1-10.

Zhu, H., D. Shang, M. Sun, S. Choi, Q. Liu, J. Hao, L. E. Figuera, F. Zhang, K. W. Choy, Y. Ao, Y. Liu, X. L. Zhang, F. Yue, M. R. Wang, L. Jin, P. I. Patel, T. Jing and X. Zhang, 2011 X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. Am J Hum Genet 88: 819-826.

Ringpfeil, F., M. G. Lebwohl, A. M. Christiano and J. Uitto, 2000 Pseudoxanthoma elasticum: mutations in the MRP6 gene encoding a transmembrane ATP-binding cassette (ABC) transporter. Proc Natl Acad Sci U S A 97: 6001-6006.

Rinn, J. L., C. Bondre, H. B. Gladstone, P. O. Brown and H. Y. Chang, 2006 Anatomic demarcation by positional variation in fibroblast gene expression programs. PLoS Genet 2: e119.

Rinn, J. L., J. K. Wang, N. Allen, S. A. Brugmann, A. J. Mikels, H. Liu, T. W. Ridky, H. S. Stadler, R. Nusse, J. A. Helms and H. Y. Chang, 2008 A dermal HOX transcriptional program regulates site-specific epidermal fate. Genes Dev 22: 303-307.

Rivas, M. A., M. Beaudoin, A. Gardet, C. Stevens, Y. Sharma, C. K. Zhang, G. Boucher, S. Ripke, D. Ellinghaus, N. Burtt, T. Fennell, A. Kirby, A. Latiano, P. Goyette, T. Green, J. Halfvarson, T. Haritunians, J. M. Korn, F. Kuruvilla, C. Lagace, B. Neale, K. S. Lo, P. Schumm, L. Torkvist, D. National Institute of, C. Digestive Kidney Diseases Inflammatory Bowel Disease Genetics, C. United Kingdom Inflammatory Bowel Disease Genetics, C. International Inflammatory Bowel Disease Genetics, M. C. Dubinsky, S. R. Brant, M. S. Silverberg, R. H. Duerr, D. Altshuler, S. Gabriel, G. Lettre, A. Franke, M. D'Amato, D. P. McGovern, J. H. Cho, J. D. Rioux, R. J. Xavier and M. J. Daly, 2011 Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43: 1066-1073.

Rivera, C. M., and B. Ren, 2013 Mapping human epigenomes. Cell 155: 39-55.

Rodriguez de Turco, E. B., W. Tang, M. K. Topham, F. Sakane, V. L. Marcheselli, C. Chen, A. Taketomi, S. M. Prescott and N. G. Bazan, 2001 Diacylglycerol kinase epsilon regulates seizure susceptibility and long-term potentiation through arachidonoyl- inositol lipid signaling. Proc Natl Acad Sci U S A 98: 4740-4745.

Rodriguez, T. A., K. E. Fernandes, K. L. Dresser and M. Duvic, 2010 Concordance rate of alopecia areata in identical twins supports both genetic and environmental factors. J Am Acad Dermatol 62: 525-527.

Roessler, E., E. Belloni, K. Gaudenz, F. Vargas, S. W. Scherer, L. C. Tsui and M. Muenke, 1997 Mutations in the C-terminal domain of Sonic Hedgehog cause holoprosencephaly. Hum Mol Genet 6: 1847-1853.

Rogers, M. A., H. Winter, L. Langbein, R. Bleiler and J. Schweizer, 2004 The human type I keratin gene family: characterization of new hair follicle specific members and evaluation of the chromosome 17q21.2 gene domain. Differentiation 72: 527-540.

Rohatgi, S., D. Clark, A. D. Kline, L. G. Jackson, J. Pie, V. Siu, F. J. Ramos, I. D. Krantz and M. A. Deardorff, 2010 Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey. Am J Med Genet A 152A: 1641-1653.

Rubio, E. D., D. J. Reiss, P. L. Welcsh, C. M. Disteche, G. N. Filippova, N. S. Baliga, R. Aebersold, J. A. Ranish and A. Krumm, 2008 CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105: 8309-8314.

Saini, V., C. D. Hose, A. Monks, K. Nagashima, B. Han, D. L. Newton, A. Millione, J. Shah, M. G. Hollingshead, K. M. Hite, M. W. Burkett, R. M. Delosh, T. E. Silvers, D. A. Scudiero and R. H. Shoemaker, 2012 Identification of CBX3 and ABCA5 as putative biomarkers for tumor stem cells in osteosarcoma. PLoS One 7: e41401.

Sakurai, D., J. Zhao, Y. Deng, J. A. Kelly, E. E. Brown, J. B. Harley, S. C. Bae, M. E. Alarcomicronn-Riquelme, J. C. Edberg, R. P. Kimberly, R. Ramsey-Goldman, M. A. Petri, J. D. Reveille, L. M. Vila, G. S. Alarcon, K. M. Kaufman, T. J. Vyse, C. O. Jacob, P. M. Gaffney, K. M. Sivils, J. A. James, D. L. Kamen, G. S. Gilkeson, T. B. Niewold, J. T. Merrill, R. H. Scofield, L. A. Criswell, A. M. Stevens, S. A. Boackle, J. H. Kim, J. Choi, B. A. Pons-Estel, B. I. Freedman, J. M. Anaya, J. Martin, C. Y. Yu, D. M. Chang, Y. W. Song, C. D. Langefeld, W. Chen, J. M. Grossman, R. M. Cantor, B. H. Hahn and B. P. Tsao, 2013 Preferential binding to Elk-1 by SLE-associated IL10 risk allele upregulates IL10 expression. PLoS Genet 9: e1003870.

Schlake, T., M. Schorpp, A. Maul-Pavicic, A. M. Malashenko and T. Boehm, 2000 Forkhead/winged-helix transcription factor Whn regulates hair keratin gene expression: molecular analysis of the nude skin phenotype. Dev Dyn 217: 368-376.

Schmidt-Ullrich, R., and R. Paus, 2005 Molecular principles of hair follicle induction and morphogenesis. Bioessays 27: 247-261.

Schmidt, A. M., T. Zou, R. P. Joshi, T. M. Leichner, M. A. Pimentel, C. L. Sommers and T. Kambayashi, 2013 Diacylglycerol kinase zeta limits the generation of natural regulatory T cells. Sci Signal 6: ra101.

Schulz, R., C. Proudhon, T. H. Bestor, K. Woodfine, C. S. Lin, S. P. Lin, M. Prissette, R. J. Oakey and D. Bourc'his, 2010 The parental non-equivalence of imprinting control regions during mammalian development and evolution. PLoS Genet 6: e1001214.

Schwanstecher, M., C. Sieverding, H. Dorschner, I. Gross, L. Aguilar-Bryan, C. Schwanstecher and J. Bryan, 1998 Potassium channel openers require ATP to bind to and act through sulfonylurea receptors. EMBO J 17: 5529-5535.

Sekimata, M., M. Perez-Melgosa, S. A. Miller, A. S. Weinmann, P. J. Sabo, R. Sandstrom, M. O. Dorschner, J. A. Stamatoyannopoulos and C. B. Wilson, 2009 CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. Immunity 31: 551-564.

Sennett, R., and M. Rendl, 2012 Mesenchymal-epithelial interactions during hair follicle morphogenesis and cycling. Semin Cell Dev Biol 23: 917-927.

Sharaf, N., M. J. Nicklin and F. S. di Giovine, 2014 Long-range DNA interactions at the IL-1/IL-36/IL-37 gene cluster (2q13) are induced by activation of monocytes. Cytokine 68: 16-22.

Sharov, A. A., T. Y. Sharova, A. N. Mardaryev, A. Tommasi di Vignano, R. Atoyan, L. Weiner, S. Yang, J. L. Brissette, G. P. Dotto and V. A. Botchkarev, 2006 Bone morphogenetic protein signaling regulates the size

of hair follicles and modulates the expression of cell cycle-associated genes. Proc Natl Acad Sci U S A 103: 18166-18171.

Shimomura, Y., and A. M. Christiano, 2010 Biology and genetics of hair. Annu Rev Genomics Hum Genet 11: 109-132.

Shimomura, Y., D. Agalliu, A. Vonica, V. Luria, M. Wajid, A. Baumer, S. Belli, L. Petukhova, A. Schinzel, A. H. Brivanlou, B. A. Barres and A. M. Christiano, 2010 APCDD1 is a novel Wnt inhibitor mutated in hereditary hypotrichosis simplex. Nature 464: 1043-1047.

Shimomura, Y., D. Agalliu, A. Vonica, V. Luria, M. Wajid, A. Baumer, S. Belli, L. Petukhova, A. Schinzel, A. H. Brivanlou, B. A. Barres and A. M. Christiano, 2010a APCDD1 is a novel Wnt inhibitor mutated in hereditary hypotrichosis simplex. Nature 464: 1043-1047.

Shimomura, Y., M. Wajid, Y. Ishii, L. Shapiro, L. Petukhova, D. Gordon and A. M. Christiano, 2008 Disruption of P2RY5, an orphan G protein-coupled receptor, underlies autosomal recessive woolly hair. Nat Genet 40: 335-339.

Shimomura, Y., M. C. Garzon, L. Kristal, L. Shapiro and A. M. Christiano, 2009a Autosomal recessive woolly hair with hypotrichosis caused by a novel homozygous mutation in the P2RY5 gene. Exp Dermatol 18: 218-221.

Shimomura, Y., M. Wajid, L. Petukhova, L. Shapiro and A. M. Christiano, 2009 Mutations in the lipase H gene underlie autosomal recessive woolly hair/hypotrichosis. J Invest Dermatol 129: 622-628.

Shimomura, Y., M. Wajid, L. Petukhova, L. Shapiro and A. M. Christiano, 2009b Mutations in the lipase H gene underlie autosomal recessive woolly hair/hypotrichosis. J Invest Dermatol 129: 622-628.

Sonoda, H., J. Aoki, T. Hiramatsu, M. Ishida, K. Bandoh, Y. Nagai, R. Taguchi, K. Inoue and H. Arai, 2002 A novel phosphatidic acid-selective phospholipase A1 that produces lysophosphatidic acid. J Biol Chem 277: 34254-34263.

Shimomura, Y., M. Wajid, L. Petukhova, M. Kurban and A. M. Christiano, 2010b Autosomal-dominant woolly hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture. Am J Hum Genet 86: 632-638.

Sun, M., N. Li, W. Dong, Z. Chen, Q. Liu, Y. Xu, G. He, Y. Shi, X. Li, J. Hao, Y. Luo, D. Shang, D. Lv, F. Ma, D. Zhang, R. Hua, C. Lu, Y. Wen, L. Cao, A. D. Irvine, W. H. McLean, Q. Dong, M. R. Wang, J. Yu, L. He, W. H. Lo and X. Zhang, 2009 Copy-number mutations on chromosome 17q24.2-q24.3 in congenital generalized hypertrichosis terminalis with or without gingival hyperplasia. Am J Hum Genet 84: 807-813.

Simon, J. M., P. G. Giresi, I. J. Davis and J. D. Lieb, 2012 Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat Protoc 7: 256-267.

Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel and W. de Laat, 2006 Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38: 1348-1354.

Skelly, D. A., M. Johansson, J. Madeoy, J. Wakefield and J. M. Akey, 2011 A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res 21: 1728-1737.

Smahi, A., G. Courtois, P. Vabres, S. Yamaoka, S. Heuertz, A. Munnich, A. Israel, N. S. Heiss, S. M. Klauck, P. Kioschis, S. Wiemann, A. Poustka, T. Esposito, T. Bardaro, F. Gianfrancesco, A. Ciccodicola, M. D'Urso, H. Woffendin, T. Jakins, D. Donnai, H. Stewart, S. J. Kenwrick, S. Aradhya, T. Yamagata, M. Levy, R. A. Lewis and D. L. Nelson, 2000 Genomic rearrangement in NEMO impairs NF-kappaB activation and is a cause of incontinentia pigmenti. The International Incontinentia Pigmenti (IP) Consortium. Nature 405: 466-472.

Smith, G. R., 2008 Meeting DNA palindromes head-to-head. Genes Dev 22: 2612-2620.

Sobreira, N. L., E. T. Cirulli, D. Avramopoulos, E. Wohler, G. L. Oswald, E. L. Stevens, D. Ge, K. V. Shianna, J. P. Smith, J. M. Maia, C. E. Gumbs, J. Pevsner, G. Thomas, D. Valle, J. E. Hoover-Fong and D. B. Goldstein, 2010 Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6: e1000991.

Song, H. H., W. Shi, Y. Y. Xiang and J. Filmus, 2005 The loss of glypican-3 induces alterations in Wnt signaling. J Biol Chem 280: 2116-2125.

Strid, J., S. J. Roberts, R. B. Filler, J. M. Lewis, B. Y. Kwong, W. Schpero, D. H. Kaplan, A. C. Hayday and M. Girardi, 2008 Acute upregulation of an NKG2D ligand promotes rapid reorganization of a local immune compartment with pleiotropic effects on carcinogenesis. Nat Immunol 9: 146-154.

Sun, H., R. S. Molday and J. Nathans, 1999 Retinal stimulates ATP hydrolysis by purified and reconstituted ABCR, the photoreceptor-specific ATP-binding cassette transporter responsible for Stargardt disease. J Biol Chem 274: 8269-8281.

Vahedi, G., Y. Kanno, Y. Furumoto, K. Jiang, S. C. Parker, M. R. Erdos, S. R. Davis, R. Roychoudhuri, N. P. Restifo, M. Gadina, Z. Tang, Y. Ruan, F. S.

Collins, V. Sartorelli and J. J. O'Shea, 2015 Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature.

van der Steen, P., H. Traupe, R. Happle, J. Boezeman, R. Strater and H. Hamm, 1992 The genetic risk for alopecia areata in first degree relatives of severely affected patients. An estimate. Acta Derm Venereol 72: 373-375.

Verlaan, D. J., S. Berlivet, G. M. Hunninghake, A. M. Madore, M. Lariviere, S. Moussette, E. Grundberg, T. Kwan, M. Ouimet, B. Ge, R. Hoberman, M. Swiatek, J. Dias, K. C. Lam, V. Koka, E. Harmsen, M. Soto-Quiros, L. Avila, J. C. Celedon, S. T. Weiss, K. Dewar, D. Sinnett, C. Laprise, B. A. Raby, T. Pastinen and A. K. Naumova, 2009 Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease. Am J Hum Genet 85: 377-393.

Williams, A., G. R. Lee, C. G. Spilianakis, S. S. Hwang, S. C. Eisenbarth and R. A. Flavell, 2013 Hypersensitive site 6 of the Th2 locus control region is essential for Th2 cytokine expression. Proc Natl Acad Sci U S A 110: 6955-6960.

Xing, L., Z. Dai, A. Jabbari, J. E. Cerise, C. A. Higgins, W. Gong, A. de Jong, S. Harel, G. M. DeStefano, L. Rothman, P. Singh, L. Petukhova, J. Mackay-Wiggan, A. M. Christiano and R. Clynes, 2014 Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition. Nat Med 20: 1043-1049.

Sundberg, J. P., K. A. Silva, R. Li, G. A. Cox and L. E. King, 2004 Adult-onset Alopecia areata is a complex polygenic trait in the C3H/HeJ mouse model. J Invest Dermatol 123: 294-297.

Sutherland, C. L., N. J. Chalupny, K. Schooley, T. VandenBos, M. Kubin and D. Cosman, 2002 UL16-binding proteins, novel MHC class I-related proteins, bind to NKG2D and activate multiple signaling pathways in primary NK cells. J Immunol 168: 671-679.

Sutton, E., J. Hughes, S. White, R. Sekido, J. Tan, V. Arboleda, N. Rogers, K. Knower, L. Rowley, H. Eyre, K. Rizzoti, D. McAninch, J. Goncalves, J. Slee, E. Turbitt, D. Bruno, H. Bengtsson, V. Harley, E. Vilain, A. Sinclair, R. Lovell-Badge and P. Thomas, 2011 Identification of SOX3 as an XX male sex reversal gene in mice and humans. J Clin Invest 121: 328-341.

Tadin, M., E. Braverman, S. Cianfarani, A. J. Sobrino, B. Levy, A. M. Christiano and D. Warburton, 2001 Complex cytogenetic rearrangement of chromosome 8q in a case of Ambras syndrome. Am J Med Genet 102: 100-104.

Tarling, E. J., T. Q. de Aguiar Vallim and P. A. Edwards, 2013 Role of ABC transporters in lipid transport and human disease. Trends Endocrinol Metab 24: 342-350.

Thomas, N. S., J. Chelly, J. Zonana, K. J. Davies, S. Morgan, J. Gault, K. A. Rack, V. J. Buckle, N. Brockdorff, A. Clarke and et al., 1993 Characterisation of molecular DNA rearrangements within the Xq12-q13.1 region, in three patients with X-linked hypohidrotic ectodermal dysplasia (EDA). Hum Mol Genet 2: 1679-1685.

Torgyekes, E., A. L. Shanske, K. Anyane-Yeboa, O. Nahum, S. Pirzadeh, E. Blumfield, V. Jobanputra, D. Warburton and B. Levy, 2011 The proximal chromosome 14q microdeletion syndrome: delineation of the phenotype using high resolution SNP oligonucleotide microarray analysis (SOMA) and review of the literature. Am J Med Genet A 155A: 1884-1896.

Tran, T. H., A. Jarrell, G. E. Zentner, A. Welsh, I. Brownell, P. C. Scacheri and R. Atit, 2010 Role of canonical Wnt signaling/ss-catenin via Dermo1 in cranial dermal cell development. Development 137: 3973-3984.

Tsurusaki, Y., N. Okamoto, H. Ohashi, T. Kosho, Y. Imai, Y. Hibi-Ko, T. Kaname, K. Naritomi, H. Kawame, K. Wakui, Y. Fukushima, T. Homma, M. Kato, Y. Hiraki, T. Yamagata, S. Yano, S. Mizuno, S. Sakazume, T. Ishii, T. Nagai, M. Shiina, K. Ogata, T. Ohta, N. Niikawa, S. Miyatake, I. Okada, T. Mizuguchi, H. Doi, H. Saitsu, N. Miyake and N. Matsumoto, 2012 Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat Genet 44: 376-378.

van Blitterswijk, W. J., and B. Houssa, 2000 Properties and functions of diacylglycerol kinases. Cell Signal 12: 595-605.

van Bon, B. W., C. Gilissen, D. K. Grange, R. C. Hennekam, H. Kayserili, H. Engels, H. Reutter, J. R. Ostergaard, E. Morava, K. Tsiakas, B. Isidor, M. Le Merrer, M. Eser, N. Wieskamp, P. de Vries, M. Steehouwer, J. A. Veltman, S. P. Robertson, H. G. Brunner, B. B. de Vries and A. Hoischen, 2012 Cantu syndrome is caused by mutations in ABCC9. Am J Hum Genet 90: 1094-1101.

Van Hemel, J. O., and H. J. Eussen, 2000 Interchromosomal insertions. Identification of five cases and a review. Hum Genet 107: 415-432.

van Steensel, B., and S. Henikoff, 2000 Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18: 424-428.

Velagaleti, G. V., S. M. Jalal, R. C. Michaelis, T. F. Rowe, J. R. Nichols and L. H. Lockhart, 2003 Molecular cytogenetic characterization of a de novo unbalanced translocation leading to trisomy 17q25-->qter and monosomy 18p11.3-->pter in a girl with dysmorphic features. Clin Dysmorphol 12: 29-33.

Visel, A., S. Minovitsky, I. Dubchak and L. A. Pennacchio, 2007 VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35: D88-92.

Visser, M., R. J. Palstra and M. Kayser, 2015 Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. Hum Mol Genet.

Vsevolodov, E. B., K. D. Ochilov, A. P. Vorob'evskii and V. M. Fedoseenko, 1983 [Cellular interrelationships of the outer root sheath of the hair follicle]. Arkh Anat Gistol Embriol 84: 64-69.

Wang, K., S. P. Dickson, C. A. Stolle, I. D. Krantz, D. B. Goldstein and H. Hakonarson, 2010 Interpretation of association signals and identification of causal variants from genome-wide association studies. Am J Hum Genet 86: 730-742.

Wei, X., V. Walia, J. C. Lin, J. K. Teer, T. D. Prickett, J. Gartner, S. Davis, N. C. S. Program, K. Stemke-Hale, M. A. Davies, J. E. Gershenwald, W. Robinson, S. Robinson, S. A. Rosenberg and Y. Samuels, 2011 Exome sequencing identifies GRIN2A as frequently mutated in melanoma. Nat Genet 43: 442-446.

Weiler, K. S., and B. T. Wakimoto, 1995 Heterochromatin and gene expression in Drosophila. Annu Rev Genet 29: 577-605.

Wieland, G. D., N. Nehmann, D. Muller, H. Eibel, U. Siebenlist, J. Suhnel, P. F. Zipfel and C. Skerka, 2005 Early growth response proteins EGR-4 and EGR-3 interact with immune inflammatory mediators NF-kappaB p50 and p65. J Cell Sci 118: 3203-3212.

Winter, H., M. A. Rogers, L. Langbein, H. P. Stevens, I. M. Leigh, C. Labreze, S. Roul, A. Taieb, T. Krieg and J. Schweizer, 1997 Mutations in the hair cortex keratin hHb6 cause the inherited hair disease monilethrix. Nat Genet 16: 372-374.

Woods, K. S., M. Cundall, J. Turton, K. Rizotti, A. Mehta, R. Palmer, J. Wong, W. K. Chong, M. Al-Zyoud, M. El-Ali, T. Otonkoski, J. P. Martinez-Barbera, P. Q. Thomas, I. C. Robinson, R. Lovell-Badge, K. J. Woodward and M. T. Dattani, 2005 Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism. Am J Hum Genet 76: 833-849.

Wu, Q. F., L. Yang, S. Li, Q. Wang, X. B. Yuan, X. Gao, L. Bao and X. Zhang, 2012 Fibroblast growth factor 13 is a microtubule-stabilizing protein regulating neuronal polarization and migration. Cell 149: 1549-1564.

Xu, X., Y. Hou, X. Yin, L. Bao, A. Tang, L. Song, F. Li, S. Tsang, K. Wu, H. Wu, W. He, L. Zeng, M. Xing, R. Wu, H. Jiang, X. Liu, D. Cao, G. Guo, X. Hu, Y. Gui, Z. Li, W. Xie, X. Sun, M. Shi, Z. Cai, B. Wang, M. Zhong, J. Li, Z. Lu, N.

Gu, X. Zhang, L. Goodman, L. Bolund, J. Wang, H. Yang, K. Kristiansen, M. Dean, Y. Li and J. Wang, 2012 Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148: 886-895.

Xu, Z., and J. A. Taylor, 2009 SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res 37: W600-605.

Yamanishi, K., 1996 Gene analysis of human skin and skin diseases. J Dermatol Sci 11: 169-176.

Yamauchi, K., and A. Kurosaka, 2010 Expression and function of glycogen synthase kinase-3 in human hair follicles. Arch Dermatol Res 302: 263-270.

Ye, D., I. Meurs, M. Ohigashi, L. Calpe-Berdiel, K. L. Habets, Y. Zhao, Y. Kubo, A. Yamaguchi, T. J. Van Berkel, T. Nishi and M. Van Eck, 2010 Macrophage ABCA5 deficiency influences cellular cholesterol efflux and increases susceptibility to atherosclerosis in female LDLr knockout mice. Biochem Biophys Res Commun 395: 387-394.

Zhao, Z., G. Tavoosidana, M. Sjolinder, A. Gondor, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K. S. Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti and R. Ohlsson, 2006 Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38: 1341-1347.

Zhu, H., D. Shang, M. Sun, S. Choi, Q. Liu, J. Hao, L. E. Figuera, F. Zhang, K. W. Choy, Y. Ao, Y. Liu, X. L. Zhang, F. Yue, M. R. Wang, L. Jin, P. I. Patel, T. Jing and X. Zhang, 2011a X-linked congenital hypertrichosis syndrome is associated with interchromosomal insertions mediated by a human-specific palindrome near SOX3. Am J Hum Genet 88: 819-826.

Zhu, Q., D. Ge, J. M. Maia, M. Zhu, S. Petrovski, S. P. Dickson, E. L. Heinzen, K. V. Shianna and D. B. Goldstein, 2011b A genome-wide comparison of the functional properties of rare and common genetic variants in humans. Am J Hum Genet 88: 458-468.

Zuk, O., S. F. Schaffner, K. Samocha, R. Do, E. Hechter, S. Kathiresan, M. J. Daly, B. M. Neale, S. R. Sunyaev and E. S. Lander, 2014 Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A 111: E455-464.

Zwijnenburg, P. J., H. Meijers-Heijboer and D. I. Boomsma, 2010 Identical but not the same: the value of discordant monozygotic twins in genetic research. Am J Med Genet B Neuropsychiatr Genet 153B: 1134-1149.